UNIVERSITAT DE VALÈNCIA

DOCTORAL THESIS

July 2021

Doctoral Programme in Quantitative Finance and Economy

TWO RELEVANT FORECASTING PROBLEMS FOR PRACTITIONERS IN FINANCE: EQUITY RISK PREMIUM AND NON-PERFORMING LOANS

David Cortés Sánchez

Thesis supervisor:

Pilar Soriano Felipe

Acknowledgements

First of all, I would like to thank Dr Pilar Soriano Felipe (my supervisor). She guided me through all this long road and advised, corrected, suggested, and enlightened me on those difficult days on which one wants to quit everything. Her contributions have increased the quality of the job significantly.

I am also very thankful to professor Alfonso Novales, who encouraged me to participate in the doctoral program, and professor Angel Pardo who has followed our progress during the doctoral programme, contributed to our research, and gave us all the needed support.

I am also grateful for suggestions, comments, and contributions from Dr Ricardo Laborda, professor Juán Alberto Sanchís, Félix Lores and Dr Alfonso Ugarte. Equally, I would like to thank participants at MAF2018, BBVA Community of Practice, colleagues at the Universitat de València and several anonymous referees for many comments that significantly improved the quality of my research.

I would also like to express my sincere gratitude to the organization and members of the Doctoral Programme in Quantitative Finance and Economy, and the financial support from the Spanish Ministry of Science and Innovation and FEDER funds (Project PGC2018–093645-B-100). Sponsors were not involved in the project.

Finally, I thank my wife Clara and my sons Pablo and Guillermo for their support and patience all those missing weekends and nights at which “I had to study”. Moreover, to my parents, grandparents, sister, aunts, uncles, cousins, and friends who have always encouraged me in everything I did and have been there when necessary and when not. I would also like to mention Dr Joaquina Paricio and professor Jose Antonio Martínez, who guided me at the right time.

Contents

Tables and Figures

Chapter 1

Can we forecast the equity risk premium in the European Monetary Union?

Chapter 2

Forecasting the European Monetary Union equity risk premium with regression trees

Chapter 3

Macro determinants of non-performing loans: a comparative panel analysis between consumer and mortgage loans

Resumen amplio

Tanto los profesionales del sector financiero como los académicos, a menudo analizan una gran cantidad de variables económicas y no económicas para formular predicciones de los activos financieros, y/o desarrollar modelos de valoración. En este sentido, los modelos de factores proporcionan una herramienta valiosa para proceder de esta manera y obtener valoraciones y predicciones a partir de un grupo de variables que muestran una elevada relación con el activo o activos a estudiar. Entre estos factores, es habitual encontrar variables económicas y financieras, que tratan de sintetizar la evolución de la actividad y las políticas económicas, factores fundamentales, mucho más vinculados a características específicas de los activos y su desempeño, o incluso factores técnicos, que analizando la evolución de los precios y los volúmenes negociados del activo, tratan de identificar cuál está siendo el sentimiento de los inversores.

Esta tesis investiga esta forma práctica de predecir el desempeño futuro de los activos financieros centrándose en dos activos muy analizados por los profesionales del sector bancario: la renta variable y el crédito a los hogares. En los dos primeros capítulos, examinamos un conjunto de variables económicas y factores técnicos para comprobar si pueden pronosticar los mercados de renta variable. En el tercer capítulo, también se exploran factores económicos similares pero, en este caso, para predecir la tasa de mora de los créditos hipotecarios y al consumo.

El objetivo principal es analizar si los factores seleccionados muestran capacidad para predecir tanto "dentro" como "fuera de muestra" y si, además, estas predicciones generan valor económico tanto a inversores como a cualquier otro agente económico interesado en predecir la evolución de estos dos activos estudiados. En este sentido, pensamos que el trabajo realizado podría contribuir a la literatura de tres maneras. Primero, expandiendo metodologías y predictores muy estudiados por la literatura a nuevas muestras de datos que permiten aportar nueva evidencia. Segundo, introduciendo otras metodologías estadísticas menos utilizadas, como pueden ser los árboles de regresión. Tercero, investigando si al utilizar muestras de datos más reducidas, como las que en ocasiones han de analizar los profesionales del sector financiero, se obtienen resultados parecidos a los obtenidos en otros estudios similares.

Respecto al primer activo en el que se centra esta tesis, la literatura económica ha estudiado decenas de variables y propuesto múltiples modelos econométricos para examinar si existe evidencia de que los rendimientos de las acciones son predecibles. Los dos primeros capítulos de la tesis, titulados "Can we forecast the equity risk premium in the European Monetary Union?" y "Forecasting the European Monetary Union equity risk premium with regression trees", se centran en la predicción de los excesos de rendimiento de las acciones a partir de múltiples predictores y amplían la investigación ya existente de dos maneras. Primero, explorando si los resultados obtenidos por la literatura son extrapolables a una muestra de datos más reducida: la Unión Monetaria Europea. Segundo, mediante la introducción de técnicas de "machine learning", como son los algoritmos de árboles de regresión, para examinar si estos enfoques proporcionan resultados de predicción superiores.

El tercer capítulo cambia de activo, y analiza la capacidad de los factores macroeconómicos para predecir la evolución de las tasas de morosidad de los hogares. Desde la gran crisis financiera de 2007-2008, las causas subyacentes de los préstamos en mora y su relación con los ciclos económicos son un tema de estudio muy relevante para los responsables de desarrollar las políticas económicas, así como para los profesionales del sector bancario. La literatura existente ha evidenciado repetidamente una clara relación inversa entre el crecimiento económico y los incumplimientos de los préstamos.

El tercer capítulo, titulado "Macro determinants of non-performing loans: a comparative panel analysis between consumer and mortgage loans", tiene como principal objetivo comprobar si algunos de los factores macroeconómicos que se han probado como relevantes para predecir la evolución de la morosidad del sistema financiero, también lo son cuando trabajamos a nivel más desagregado, y con muestras de datos más reducidas. En este caso, analizamos una cartera de préstamos hipotecarios, y otra de préstamos al consumo, mediante un panel no balanceado que incluye los siguientes ocho países: Estados Unidos, España, México, Turquía, Colombia, Perú, Argentina y Chile.

Cada uno de los capítulos que componen la tesis está explicados en las diferentes secciones que siguen.

R.1 Can we forecast the equity risk premium in the European Monetary Union?

Este capítulo investiga la capacidad de múltiples variables económicas e indicadores técnicos para predecir la prima de riesgo del índice de renta variable MSCI EMU, tanto en estimaciones "dentro de muestra" como "fuera de muestra". Además, también estudia si estos mismos predictores pueden generar valor económico para un inversor que es averso al riesgo e invierte en una cartera formada por dos tipos de activos: renta variable (MSCI EMU) y/o mercado monetario (Euribor a 1 mes).

Predecir la evolución de los mercados de acciones es uno de los temas más populares tanto para los académicos como para los profesionales de las finanzas. La literatura existente ha estudiado numerosas variables y propuesto múltiples modelos econométricos para esclarecer si existe evidencia empírica de la capacidad para predecir los rendimientos de los mercados de renta variable. En general, muchos economistas aceptan que los rendimientos de las acciones sí son predecibles en ejercicios que se realizan "dentro de muestra", pero este mismo consenso no existe cuando los ejercicios de predicción se realizan "fuera de muestra".

El capítulo, siguiendo la línea de investigación iniciada por Neely et al. (2014), analiza la capacidad predictiva de un amplio conjunto de variables económicas e indicadores técnicos en los mercados de renta variable de la Unión Monetaria Europea. Primero, analiza la capacidad predictiva de las variables "dentro de muestra" utilizando modelos de regresión lineal simple, y modelos de regresión múltiples en los que se incorpora un análisis de componentes principales. Posteriormente, realiza otro ejercicio de predicción "fuera de muestra", para confirmar los resultados obtenidos previamente.

Para saber si las variables explicativas incluidas en los modelos tienen capacidad predictiva "fuera de muestra", se comparan los errores cuadráticos medios de las predicciones obtenidas, con los errores cuadráticos medios del promedio histórico de los rendimientos en exceso. De esta forma, si los errores obtenidos con los predictores son inferiores, se considera que éstos tienen capacidad predictiva.

Finalmente, para concluir el análisis, también se estudia si las predicciones obtenidas "fuera de muestra" permiten generar valor económico para un inversor con diferentes grados

de aversión al riesgo, una función de utilidad cuadrática y que invierte únicamente en dos tipologías de activos: renta variable y/o el activo libre de riesgo.

R.1.1 Muestra

Los datos incluyen información mensual desde enero de 2000 hasta diciembre de 2020. Las fuentes son: Bloomberg, Institutional Brokers' Estimate System (I/B/E/S) a partir de Refinitiv, y Haver Analystics.

La variable dependiente, la prima de riesgo de las acciones de la zona euro, se calcula sustrayendo a la tasa de variación mensual del índice de renta variable MSCI EMU la rentabilidad del Euribor a un mes (Euribor 1M). Como predictores se seleccionan doce variables que recogen factores fundamentales y económicos, y catorce indicadores técnicos obtenidos a partir de indicadores de medias móviles y tendencias de los precios de las acciones, así como medidas de volúmenes negociados.

R.1.2 Metodología

El marco convencional para analizar la predictibilidad de la prima de riesgo de las acciones se basa en regresiones lineales. Los excesos de retorno de la renta variables en el periodo t+1 se estiman a partir del comportamiento de uno o múltiples predictores en el periodo anterior t.

Las estimaciones se realizan tanto con modelos lineales simples, en los que se incluyen un único predictor, como con regresiones lineales múltiples, en las que se incluyen como variables explicativas componentes principales calculadas a partir de todos los predictores.

La estimación de las regresiones se realiza por mínimos cuadrados ordinarios (MCO), lo que puede acarrear una serie de problemas estadísticos como son la correlación serial y/o la heteroscedasticidad de los estimadores, así como el sesgo de Stambaugh (1986,1999). Para solucionar el primer problema, se utilizan los errores estándar de Newey y West (1987). Mientras que para eliminar el sesgo de Stambaugh, se generan con técnicas de remuestreo unos nuevos "p-values" que sí tienen en cuenta la persistencia de los regresores y las correlaciones existentes entre la variable explicativa retardada y la endógena.

La capacidad predictiva de los factores "fuera de muestra" se calcula dividiendo la muestra en dos subconjuntos. Uno, denominado ventana de estimación, y que inicialmente se extiende desde enero del 2000 hasta diciembre del 2012, sirve para realizar las estimaciones de los parámetros de la ecuación. El otro, denominado ventana de predicción y que incluye de enero 2013 hasta diciembre de 2020, es sobre el que se realizan las predicciones. Las predicciones se calculan mensualmente, y una vez que se estima el exceso de rendimientos de un periodo, los datos reales de este periodo se incluyen en la ventana de estimación, de forma que esta va creciendo progresivamente.

Para saber si las variables explicativas incluidas en los modelos tienen capacidad predictiva "fuera de muestra", las predicciones obtenidas con los predictores seleccionados se comparan con las predicciones obtenidas con el promedio histórico de los excesos de retorno. El desempeño de los pronósticos se analiza en términos del R^2 de Campbell y Thompson (R^2_{os}), que compara los errores cuadráticos medios (MSFE) de las regresiones construidas con los predictores seleccionados con los errores cuadráticos medios del predictor de referencia, en este caso el promedio histórico de los excesos de retorno. Si el $R^2_{os} > 0$, entonces asumimos que el factor o factores seleccionados sí muestran habilidad predictiva.

Además de calcular el R^2_{os}, comprobamos si los resultados obtenidos son estadísticamente significativos, y también realizamos una descomposición de los errores cuadráticos medios tal y como propuso Theil (1971).

Como ejercicio final, y a partir de los análisis de Ferreira y Santa-Clara (2011), se estudia si las predicciones "fuera de muestra" tienen capacidad de generar valor económico. Para ello, se asume que las inversiones las realiza un inversor con diferentes grados de aversión al riesgo, que intenta maximizar su función de utilidad esperada, que es cuadrática, y que únicamente invierte en dos activos financieros: renta variable (MSCI EMU), y/o el activo libre de riesgo (Euribor 1M).

Los pesos que la cartera asigna a cada uno de los activos se obtienen a partir de los pesos óptimos calculados por Markowitz, y que resultan de las predicciones históricas calculadas por cada uno de los predictores para los excesos de retornos y su volatilidad.

Para saber si cada una de las estrategias genera o no valor económico, se compara la utilidad de cada una de las estrategias respecto a la de la cartera "benchmark", que se construye a partir de los rendimientos promedios históricos. Si la utilidad de la estrategia es superior a la

utilidad del "benchmark", entonces se considera que la estrategia genera valor económico. Puesto que la utilidad que se obtiene se puede interpretar como una aproximación del "Certainty Equivalent Return" (CER) que el inversor exige, lo que el ejercicio de asignación de activos hace es comparar el CER de cada una de las estrategias, siendo el CER más alto el que mayor utilidad proporciona al inversor.

Finalmente, junto con el CER, calculamos otros dos criterios para valorar las diferentes estrategias. Uno es el ratio de Sharpe, calculado como el cociente de los excesos de rendimiento de la cartera seleccionada y la desviación estándar de esos mismos excesos de rendimiento. Y el "Turnover Ratio", que calcula la frecuencia y magnitud con la que cada estrategia reasigna activos, y permite calcular un coste de transacción de cada una de las estrategias de inversión. Este coste disminuye el CER de cada una de las estrategias, penalizando en mayor grado aquellas que son más agresivas e implican mayores compra-ventas en su reasignación de activos.

R.1.3 Resultados y conclusiones

Los resultados obtenidos confirman algunos de los hallazgos observados por Neely et al. (2014) para el mercado de renta variable de Estados Unidos, pero contradicen otros. Entre los resultados que se confirman para la Unión Económica y Monetaria (UEM), destacan que la gran mayoría de los predictores técnicos y las regresiones múltiples con componentes principales muestran capacidad predictiva "dentro de muestra". Por el contrario, los resultados "fuera de muestra" no confirman los obtenidos "dentro de muestra". Las variables técnicas, y sus componentes principales dejan de tener capacidad predictiva, y solo unas pocas variables económicas exhiben capacidad predictiva "fuera de muestra" y crean valor económico para un inversor con un mayor grado de tolerancia al riesgo. Tan solo el factor Book-to-Market (BM) es capaz de predecir "dentro de muestra", "fuera de muestra", y crear valor económico a un inversor con cualquier nivel de aversión al riesgo.

Los resultados alcanzados son relevantes para profesionales del sector financiero en la medida que aportan más luz sobre aquellas variables económicas y técnicas que pueden ser útiles para predecir en los mercados de renta variable, y permiten comparar si similares reglas de inversión pueden funcionar en diferentes geografías. En este capítulo se confirma cómo algunos de los predictores más analizados por la literatura sí muestran capacidad predictiva

"dentro de muestra" en la UEM. Sin embargo, solo un grupo muy reducido de ellos tiene capacidad para predecir "fuera de muestra" la prima de riesgo de las acciones, y además, generar valor económico.

R.2 Forecasting the European Monetary Union equity risk premium with regression trees

El capítulo 2, amplía el análisis realizado en el primero, y estudia si los métodos basados en árboles de clasificación y regresión (CART), pueden ayudar a mejorar las predicciones "fuera de muestra" de los excesos de rendimiento de las acciones en la Unión Monetaria Europea.

Los árboles de clasificación y regresión fueron introducidos por Breiman et al. (1984), y poseen algunas ventajas que los hacen interesantes a la hora de predecir las primas de riesgo de las acciones. Una de ellas es su sencillez, tanto para implementarlos como para interpretar sus resultados. Otra es que no necesitan de supuestos. Al ser modelos no paramétricos, no requieren asumir ningún comportamiento de las variables, ni de cómo se relacionan éstas. Por tanto, estos modelos permiten recoger las posibles relaciones lineales y no lineales que existan entre las variables explicativas y las endógenas. Y otra ventaja de los árboles es que permiten trabajar con un elevado número de variables explicativas y obtener resultados consistentes.

Al igual que cualquier otro modelo estadístico, los modelos CART también muestran algunas limitaciones, y en este caso, la principal es su inestabilidad en los resultados. Esto significa que pequeños cambios en los datos de la muestra original pueden generar resultados muy diferentes. Para solucionar este problema de inestabilidad, la literatura ha desarrollado soluciones de agregación o ensamble ("ensamble methods") que tratan de disminuir esta variabilidad.

Este segundo capítulo centra su atención en los árboles de regresión, e investiga su capacidad para seleccionar predictores y realizar predicciones "fuera de muestra" de la prima de riesgo de las acciones de la zona euro. Además, investiga la capacidad que tienen estos métodos estadísticos para generar valor económico a un inversor, con aversión al riesgo, y que invierte en dos activos: acciones y/o en el activo libre de riesgo. Para ello, se maximiza la

función de utilidad de los inversores haciendo uso del método de Brandt y Santa-Clara (2006) de selección dinámica de carteras.

R.2.1 Muestra

Los datos analizados en el capítulo son de frecuencia mensual, incluyen información desde enero de 2000 hasta diciembre de 2020, y se obtienen de las bases de datos de Bloomberg, Institutional Brokers' Estimate System (I/B/E/S) a partir de Refinitiv, y Haver Analytics.

La variable dependiente, la prima de riesgo de las acciones de la zona euro, se calcula sustrayendo a la tasa de variación mensual del índice de renta variable MSCI EMU la rentabilidad del Euribor a un mes (Euribor 1M). Como predictores se seleccionan 26 variables muy utilizadas tanto por académicos como por profesionales de las finanzas, y que incluyen factores fundamentales como el PER o los beneficios por acción, variables económico-financieras, y factores técnicos que tratan de captar el sentimiento de los inversores. Las variables se analizan tanto en niveles, como en primeras diferencias.

R.2.2 Metodología

Para analizar la predictibilidad de la prima de riesgo de las acciones se predicen los excesos de retorno en el periodo t+1 a partir del comportamiento de uno o múltiples predictores en el periodo anterior t. Las predicciones se realizan "fuera de muestra", y haciendo uso de tres técnicas de árboles de regresión ("bagging", "random forests" y "boosting"), de modelos de regresión lineal simple con cada uno de los predictores, y a partir de los promedios históricos de los excesos de rendimiento de las acciones, siendo estas últimas predicciones las que se utilizan como referencia para comparar la capacidad predictiva de cada uno de los modelos. De nuevo, y como en el capítulo primero, el poder predictivo de las variables y los modelos se estudia comparando los errores cuadráticos medios de todas las estrategias de predicción con el R^2 de Campbell y Thompson, y realizando la descomposición de Theil (1971) de los errores obtenidos fuera de muestra.

Los modelos CART fueron introducidos por Breiman, Friedman, Olshen y Stone en 1984, en su libro "Classification and Regression Trees". Se trata de una técnica no paramétrica, que permite identificar qué variables explicativas ayudan a predecir mejor la variable dependiente, dando como resultado un árbol de decisión. La idea principal es dividir recursivamente el

espacio de datos en sub-espacios más reducidos en los que se agrupan valores-respuesta similares. Una vez completada la separación, se predice un valor constante de la variable-respuesta dentro de cada área. La principal diferencia entre los árboles de clasificación y de regresión es que, en el primer caso, la variable dependiente es de tipo categórica, mientras que en el segundo, la variable dependiente es continua.

Uno de los principales problemas que presentan los árboles de clasificación y regresión es la elevada inestabilidad de sus resultados ante pequeños cambios en la muestra. Esto ocurre porque en la partición recursiva que realizan los árboles, la variable seleccionada y el punto de corte exacto seleccionado en cada nodo determina cómo se dividen las observaciones en los nodos subsiguientes. Por esta razón, los árboles son muy inestables y su estructura puede cambiar drásticamente si las primeras variables seleccionadas y puntos de corte cambian ante pequeños cambios en la muestra.

Para solucionar estos problemas de inestabilidad, la literatura ha desarrollado soluciones que se centran en generar múltiples árboles, y ponderar o agregar los resultados obtenidos, de tal forma que la variabilidad se reduce significativamente. En este capítulo se utilizan tres de estos modelos, también conocidos como modelos de "ensamble". Los dos primeros, "bagging" y "random forests" fueron incialmente propuestos por Breiman (1996a, 1996b y 2001), y en ellos se generan múltiples árboles de decisión mediante técnicas de "bootstrapping", y los resultados finales se obtienen del promedio de todos ellos. Por otro lado, Freund y Schapire (1997), propusieron un modelo iterativo, conocido como "boosting", en el que los árboles se estiman secuencialmente, y cada árbol nuevo se ajusta a una versión modificada del conjunto de datos original.

Para finalizar el capítulo y una vez analizada la capacidad predictiva de los modelos de árboles de regresión, se estudia si los principales predictores seleccionados por éstos pueden generar valor económico para un inversor con aversión al riesgo, que tiene una función de utilidad cuadrática, y que invierte únicamente en dos activos: renta variable y/o el activo libre de riesgo.

Considerando la metodología propuesta por Brandt y Santa-Clara (2006) estimamos los pesos óptimos de la cartera a partir de las predicciones obtenidas con las variables seleccionadas, y se crean carteras dinámicas que varían sus ponderaciones según la evolución de los predictores. La ventaja del modelo propuesto por Brandt y Santa Clara (2006) es que permite asignar los pesos óptimos de la cartera en función de cuál está siendo la evolución de

uno o varios predictores a la vez. El modelo asume que las ponderaciones de las carteras son una combinación lineal de estos factores, también conocidos como variables de estado.

Teniendo en cuenta esta relación lineal y haciendo uso de los pesos óptimos calculados por Markowitz, se obtienen los pesos que la cartera del inversor asigna a cada uno de los activos. Una vez conocidos éstos, se calcula la riqueza y la utilidad generada para cada inversor por cada una de las estrategias.

Finalmente, y al igual que se hizo en el capítulo primero, se compara la utilidad de cada una de las estrategias, el CER, respecto a la de la cartera "benchmark". En esta ocasión, la cartera "benchmark" seleccionada se construye fijando de manera permanente el peso del activo de renta fija en el 75%, y el de la renta variable en el 25%. Finalmente, junto con el CER, se vuelven a calcular otros dos criterios de valoración de las estrategias: el ratio de Sharpe y el Turnover Ratio.

R.2.3 Resultados y conclusiones

Los resultados "fuera de muestra" indican que el uso de los árboles de regresión para identificar las relaciones entre las variables y realizar predicciones no aporta mayor capacidad predictiva. Los tres métodos utilizados, "bagging", "random forests" o "boosting", muestran unos errores cuadráticos medios de predicción superiores a las predicciones realizadas con los promedios históricos. Sin embargo, al realizar la descomposición de Theil, los errores de los árboles presentan un menor sesgo y varianza, y unas mayores covarianzas que los errores del "benchmark". Esto podría sugerir que para realizar comparaciones de la capacidad predictiva de diferentes modelos estadísticos, también sería conveniente comparar otros momentos de la función de densidad de las predicciones, además de sus errores cuadrados medios (MSFE).

Por su parte, el ejercicio de asignación de activos muestra que las técnicas de árboles de regresión tampoco tienen capacidad de generar valor económico para un inversor con una función de utilidad cuadrática y diferentes niveles de aversión al riesgo. El CER relativo de la cartera "benchmark" es muy superior al obtenido por cualquiera de las estrategias de inversión construidas haciendo uso de los árboles de regresión. No obstante, los resultados obtenidos con los ratios de Sharpe no confirman los obtenidos con el CER, lo que no descarta que las estrategias construidas con árboles de regresión sí puedan tener capacidad de generar valor económico.

Por tanto, los resultados obtenidos en este segundo capítulo indican que, con la muestra estudiada, los árboles de regresión no parecen aportar mayor capacidad predictiva para estimar las primas de riesgo de las acciones, ni tampoco está claro que aporten valor económico para inversores con diferentes niveles de aversión al riesgo y una función de utilidad cuadrática. Tal vez, estos resultados se podrían explicar por la reducida dimensión de la muestra seleccionada en este capítulo que, por otra parte, es la única que se puede obtener para la EMU. En trabajos futuros, se debería incluir una muestra con una dimensión muy superior para aprovechar toda la potencia analítica de estas técnicas.

R.3. Macro determinants of non-performing loans: a comparative panel analysis between consumer and mortgage loans

En el tercer capítulo se analizan algunos de los factores macroeconómicos que ayudan a explicar la evolución de la tasa de morosidad (NPL) de las carteras de crédito de los hogares. Conocer la morosidad y sus factores determinantes es una tarea fundamental tanto para las entidades financieras como para las autoridades económicas de un país o región. Por una parte, las entidades financieras estudian la evolución pasada y futura de su morosidad para calcular, entre otras cosas, sus niveles de provisiones o el coste del riesgo del crédito de sus futuras operaciones. Por otra parte, las autoridades financieras monitorizan exhaustivamente cómo evoluciona el crédito fallido en el sistema y cómo éste puede comprometer la estabilidad del conjunto del sistema financiero.

La literatura económica ha proporcionado una amplia evidencia de que el ciclo económico y los shocks financieros se encuentran entre los principales factores sistémicos que explicarían la evolución de las tasas de morosidad. En este sentido, Manz (2019) presenta una amplia revisión de la literatura que analiza los principales factores determinantes de la morosidad del sistema financiero.

Aunque la literatura que analiza las causas de la morosidad en los sistemas financieros es extensa, la disponibilidad de datos históricos de mora a nivel desagregado y en múltiples geografías es limitada. Esto ha hecho que las técnicas econométricas de panel sean las predominantes en este tipo de estudios. Los beneficios de utilizar procedimientos de datos de panel son múltiples, y tal y como enfatiza Hsiao (1986), los análisis con datos de panel se benefician de conjuntos de datos más extensos, con una mayor variabilidad, y una menor

"colinealidad". Además, los paneles permiten controlar la heterogeneidad individual no observada de las secciones transversales.

Según Pesaran (2015), los paneles se podrían dividir en tres grandes grupos dependiendo de sus supuestos sobre el número relativo de unidades transversales (N) y el número de períodos de tiempo (T). Primero, los denominados "micropaneles", que serían aquellos en los que N es grande y T es pequeña. Segundo, los "macropaneles", aquellos en los que tanto la N como la T son grandes. Finalmente, tendríamos los paneles intermedios, en los que la N es pequeña, y la T es grande.

La categoría de panel es una característica importante porque, dependiendo del tamaño de la muestra, se pueden realizar diferentes técnicas de estimación para calcular el impacto de los factores macroeconómicos en las tasas de incumplimiento. Hasta ahora, la mayor parte de la literatura relacionada ha centrado su atención en los "micropaneles" o "macropaneles", y se ha prestado menos atención a situaciones intermedias en las que T es grande y N es pequeño.

En este tercer capítulo, exploramos si haciendo uso de paneles intermedios podemos llegar a conclusiones similares a las obtenidas previamente por la literatura haciendo uso de paneles con una mayor dimensión transversal. Concretamente, estudiamos si diversos factores macroeconómicos muestran capacidad para predecir la tasa de mora de dos carteras de crédito: una hipotecaria y otra de crédito al consumo. Para ello, trabajamos con un panel no balanceado, que incluye datos trimestrales para ocho economías desarrolladas y emergentes. Seleccionamos esta muestra heterogénea y corta de países como un ejemplo real de geografías analizadas periódicamente por analistas y reguladores. A menudo, estos profesionales se ven obligados a realizar análisis y predicciones con este tipo de paneles intermedios porque obtener y agregar datos para otros países puede ser costoso, o simplemente imposible de lograr.

R.3.1 Muestra

El estudio analiza datos trimestrales de una muestra de ocho países, que incluyen a España, México, Estados Unidos, Turquía, Colombia, Perú, Chile y Argentina. Como variables endógenas se seleccionan las tasas de mora, calculadas como el cociente entre crédito en mora durante más de 90 días y el crédito total de la cartera, de las carteras de préstamos al consumo e hipotecarios. Los datos se obtienen a nivel agregado para el sistema financiero de cada uno de los bancos centrales o entidades reguladoras de cada una de las geografías analizadas. Puesto

que no todas las geografías reportan las mismas series históricas de crédito moroso, se requiere trabajar con dos paneles no balanceados, uno para la cartera de consumo y otro para la hipotecaria, en el que la dimensión temporal del panel (T) varía entre 37 y 113 observaciones, y cubre en ambos paneles el periodo que va desde marzo de1992 hasta diciembre de 2019. En resumen, dos paneles no balanceados, con una dimensión N=8 y T=37-113.

Como variables explicativas, se consideran seis factores que la literatura económica (ver por ejemplo Manz (2019)) ha demostrado tienen un impacto significativo en las tasas de morosidad del sistema financiero. Estas son: la propia variable endógena rezagada en el tiempo, el stock de crédito de periodos anteriores, el PIB real, los tipos monetarios reales, el precio de la vivienda real, y los índices de renta variable de cada país.

R.3.2 Metodología

Dada la cantidad de datos temporales que presentan los diferentes países que componen los paneles, primero se analizan ciertos problemas y características que son más propias de las series temporales que de las técnicas de panel.

Lo primero que se comprueba es si las series son estacionarias o, en caso negativo, si estas presentan relaciones de cointegración. Para ello, primero se realiza una batería de tests de raíces unitarias que incluye los test de primera generación de Levin-Lin-Chu, Breitung, Hadri, Im-Pesaran-Shin, Fisher-ADF y Fisher-PP, y el CIPS-test propuesto por Pesaran (2005), que pertenece a una segunda generación de pruebas de raíz unitaria en las que se tiene en cuenta posibles problemas de correlación transversal en los residuos. Posteriormente, y una vez comprobada la existencia de raíces unitarias en la mayoría de variables, se comprueba la presencia de relaciones de cointegración. Para ello, calculamos los test de Pedroni (1999,2004), que son una extensión de la metodología de Engle y Granger (1987) pero en estructuras de datos de panel.

Segundo, comprobamos si existen posibles relaciones no observadas entre las variables transversales, lo que en la literatura de datos de panel se conoce como “cross-section dependence” (CSD). Ésta surge cuando existe cierta estructura de correlación en el término del error de las diferentes unidades transversales, debido a la presencia de factores comunes no observables. La “cross-section dependence” elimina la totalidad, o parte de los beneficios, de operar con un panel, y conduce a estimaciones inconsistentes de los parámetros con los

métodos tradicionales de estimación por mínimos cuadrados ordinarios (MCO). Para detectar la presencia de la correlación transversal se realiza la prueba CD de Pesaran (2004), en la que se comprueba si existe correlación entre los residuos de las estimaciones individuales de cada sección transversal.

Por último, comprobamos si es preferible estimar los coeficientes de cada una de las regresiones de manera heterogénea, como si se tratara de un sistema de ecuaciones en el que los parámetros de cada ecuación se obtienen individualmente, $\beta_n \neq \beta \ \forall n$ o, por el contrario, asumimos que los parámetros de cada variable son homogéneos entre todos los datos transversales, $\beta_n = \beta \ \forall n$ (pooling assumptiom). Este debate cobra relevancia a medida que el número de observaciones temporales crece y el número de datos permite obtener estimaciones heterogéneas consistentes para cada uno de los individuos. Para determinar si es conveniente mantener el supuesto de homogeneidad ($\beta_n = \beta \ \forall n$) en los paneles, se estima individualmente cada unidad transversal por MCO, y se realiza un test de estabilidad de Chow. Si se acepta la estabilidad de todos los coeficientes, entonces se mantiene la hipótesis de homogeneidad entre los parámetros de cada unidad transversal, de lo contrario, se asume heterogeneidad en las estimaciones.

Para estimar los paneles y comprobar la capacidad predictiva "dentro de muestra" de las ecuaciones y las variables explicativas seleccionadas, estimamos los paneles con cinco métodos diferentes. Dos modelos estáticos muy empleados en la literatura de los "micropaneles" y que serían: primero, un modelo de agregación simple (pool), donde se asume que todos los parámetros, incluida la constante, se estiman de manera homogénea. Segundo, un modelo de efectos fijos con estimaciones intragrupo (FE-WG). El problema de utilizar estos modelos estáticos cuando las ecuaciones son dinámicas, es que los estimadores obtenidos pueden presentar sesgos e inconsistencias. Por lo tanto, para eliminar los posibles problemas de endogeneidad también estimamos introduciendo variables instrumentales. Concretamente estimamos los paneles con mínimos cuadrados de dos etapas (2SLS), y con mínimos cuadrados de tres etapas (3SLS). Los estimadores en tres etapas, propuestos por Zellner y Theil (1962), al estimar por mínimos cuadrados generalizados también permiten considerar posibles problemas de correlación transversal contemporánea entre los residuos. Finalmente, y con el objetivo de comparar los resultados mediante el uso de diversas técnicas econométricas también estimamos por "Dynamic Mean Groups" (DMG), una técnica introducida por Pesaran y Smith (1995), y que es más propia de "macropaneles".

Para concluir, estudiamos cómo los paneles estimados con los cinco métodos anteriormente explicados predicen fuera de muestra. Para ello, se seleccionan tres periodos "fuera de muestra" que cubren uno, tres y cinco años respectivamente, y se compara la capacidad predictiva de cada modelo respecto a un modelo naïve que predice haciendo uso de un modelo AR(1) de las tasas de morosidad en cada país. Las comparaciones entre las predicciones, se realizan de nuevo comparando los errores cuadráticos medios generados por cada modelo, de tal forma que el modelo con menores errores es el que mejor predice.

R.3.3 Resultados y conclusiones

De los análisis preliminares de las series, obtenemos que las series presentan raíces unitarias en la gran mayoría de ellas, pero que las tasas de variación interanual son estacionarias en todos los casos. Preferimos trabajar con tasas de variación interanuales y no trimestrales, por la naturaleza lenta de los procesos de morosidad, que hacen que desde que un préstamo comienza a dar problemas hasta que se reconoce que está en mora pueden transcurrir muchos meses. Además, el análisis de cointegración sugiere que no existen relaciones de estabilidad a largo plazo en ninguno de los paneles estudiados, por lo que rechazamos la hipótesis de cointegración. Esto nos lleva a seleccionar un modelo lo más sencillo posible que trabaja en diferencias interanuales con todas las variables no estacionarias, y en la que incluimos la variable endógena rezagada un periodo.

Además, el análisis preliminar también ofrece como resultado que es mejor asumir homogeneidad para la estimación de los parámetros de los paneles ($\beta_n = \beta \ \forall n$), que heterogeneidad, y que el panel de los préstamos al consumo sí presenta "cross-section dependece", pero no así el panel de los préstamos hipotecarios.

Los resultados obtenidos en las estimaciones "dentro de muestra" están en línea con los de la literatura de referencia. Primero, las tasas de morosidad muestran una naturaleza persistente, el crecimiento crediticio de periodos anteriores tiene un impacto positivo en los préstamos dudosos, y una aceleración del PIB real, los precios reales de la vivienda y los mercados de valores, junto con menores costes de financiación, conducen a menores préstamos dudosos. Segundo, las estimaciones también indican que los préstamos al consumo y las hipotecas muestran una sensibilidad muy similar a los ciclos económicos, pero que las elasticidades son algo diferentes e invitan a estimar estas dos carteras de forma independiente. Además las

estimaciones obtenidas con todas las metodologías estadísticas son similares, y nos lleva a preguntarnos si el número de observaciones temporales es lo suficientemente amplio como para eliminar el sesgo de las estimaciones obtenidas con los modelos de efectos fijos. Finalmente, las predicciones "fuera de muestra" también confirman que la mayoría de los factores económicos juegan un papel esencial en el pronóstico de los préstamos en mora.

Este tercer capítulo intenta contribuir a la extensa literatura que explora los determinantes macroeconómicos de los préstamos en mora, y encuentra resultados similares, pero centrando los esfuerzos en paneles intermedios, aquellos con T grande y N pequeña, en los que la literatura existente es escasa, y a los que los profesionales del sector financiero han de enfrentarse en algunas ocasiones.

R.4 Conclusiones

Los profesionales del sector financiero y los académicos con frecuencia analizan una gran cantidad de variables económicas y no económicas para formular predicciones de los activos financieros, y/o desarrollar modelos de valoración. En este sentido, los modelos de factores proporcionan una herramienta valiosa para proceder de esta manera y obtener valoraciones y predicciones a partir de un grupo de variables que muestran una elevada relación con el activo o activos que estemos estudiando.

Esta tesis doctoral tiene como objetivo profundizar en este tipo de análisis centrando su atención en dos tipos de activos financieros: la renta variable y el crédito. Para el primero de los activos, este trabajo dedica dos capítulos. En el primero, titulado "Can we forecast the equity risk premium in the European Monetary Union?", analizamos si aquellos factores y metodologías más estudiados por la literatura económica para predecir la prima de riesgo de las acciones, muestran también la misma capacidad predictiva en una nueva muestra de datos para la zona Euro, y para un periodo de tiempo que abarca desde enero de 2000 hasta diciembre de 2020. Los resultados obtenidos indican que una gran parte de los predictores muestran capacidad predictiva "dentro de muestra", pero muy pocos lo hacen "fuera de muestra" y/o generan valor en las decisiones de asignación de recursos.

En capítulo 2, titulado "Forecasting the European Monetary Union equity risk premium with regression trees", introducimos técnicas de "machine learning" para ver si éstas permiten mejorar la selección de los predictores y las predicciones de la prima de riesgo de las acciones

en la Eurozona. Los resultados obtenidos indican que con la muestra temporal (datos mensuales desde enero de 2000 hasta diciembre de 2020) y las 26 variables explicativas analizadas, los árboles de regresión no aportan una mayor capacidad predictiva que un modelo simple basado en las predicciones obtenidas con los promedios históricos. Además, los modelos de árboles de regresión tampoco muestran una especial habilidad seleccionando predictores que posteriormente ayuden a los inversores a construir carteras de inversión.

De los resultados obtenidos en los dos primeros trabajos, y de cara a futuras investigaciones, observamos dos caminos claros para continuar profundizando en las investigaciones emprendidas en estos dos primeros capítulos. Por un lado, introducir modelos dinámicos que permitan alterar los predictores seleccionados y los parámetros estimados según la fase del ciclo o el régimen económico imperante en cada momento. En esta línea, técnicas como los Bayessian Model Averaging (BMA), Dynamic Models Selection (DMS), o Time Varying Parameter (TVP), podrían ser interesantes a la hora de seleccionar los predictores y obtener múltiples estimaciones de los parámetros. Por otro lado, continuar investigando con nuevos modelos de "machine learning", pero incrementando significativamente la dimensionalidad de la muestra analizada.

Por último, en el tercer capítulo, titulado "Macro determinants of non-performing loans: a comparative panel analysis between consumer and mortgage loans", también se exploran factores económicos similares, pero esta vez para predecir la tasa de morosidad de los préstamos hipotecarios y al consumo. Para ello se estudia un panel no balanceado que incluye ocho países heterogéneos y una muestra temporal de datos que incluye datos desde 1992 hasta el 2019 en el caso del país que más datos aporta al panel, y desde el 2011 al 2019 en el caso que menos.

Los resultados obtenidos con el panel intermedio propuesto, coinciden con los resultados obtenidos por una gran parte de la literatura haciendo uso de paneles de mayor dimensión transversal, y confirman que los factores macroeconómicos son relevantes para predecir la tasa de mora de las carteras hipotecarias y de consumo. Estos resultados son interesantes porque son semejantes a los obtenidos por la literatura económica, pero haciendo uso de paneles con una dimensión intermedia, más similares a los que en numerosas ocasiones han de analizar los profesionales del sector financiero.

Introduction

Obtaining reliable and accurate forecasts of future asset's performance is crucial for a wide range of economic agents, including policymakers conducting economic policies, investors building portfolios and hedging multiple risks, companies making investment decisions, or even academic researchers investigating valuation models validity. A popular and extended procedure to make predictions, especially among practitioners in finance, is to use factor models. These models have existed for many years. Even before introducing the popular Sharpe's CAPM or the Ross'APT models, Markowitz had already proposed using a single factor model to explain security returns.

Factor models look at various variables, known as loading factors, which gather common and relevant information that impacts assets performance. In this sense, these techniques provide a valuable tool to assist financial analysts with the identification of pervasive factors that affect a large number of securities. Thus, these factors may include macroeconomic or financial variables that gather current economic and political conditions, fundamental variables that help identify specific asset's characteristics, and even technical variables that try to pick investors sentiment.

This thesis investigates this practical way to forecast the future assets' performance focusing on two well-studied themes in financial economics and banking: first, the ability to predict the equity risk premium, and second, the macroeconomic determinants of non-performing loans (NPL) rates. In the first two chapters, a set of economic and technical metrics are examined to check whether they can forecast equity markets. In the third chapter, similar economic factors are also explored to predict credit loan delinquencies, measured as non-performing rates.

The dissertation aims to substantiate whether macroeconomic and other relevant factors show the ability to predict equity risk premiums and loan failures, both in-sample and out-of-sample, and contribute to this rich literature in three ways. First, by expanding existing research to new datasets. Second, by introducing new econometric methodologies already used by practitioners but still growing in the related academic field. Third, by investigating whether using smaller datasets, such as those frequently used by practitioners, yield similar results to the ones with more extensive datasets within the existing academic literature. In summary, the

dissertation aims to learn from the current academic research and to contribute to it from a practitioners' perspective.

Regarding the first asset on which this thesis focuses, forecasting stock returns is probably one of the financial topics that raises the most interest among financial researchers. The available literature has studied many types of variables and proposed multiple econometric models to examine whether there is evidence of returns predictability. The first two essays of this dissertation expand on the previous research in two ways. The first one, by exploring whether results obtained for long US data sets can be confirmed in the European Monetary Union, where the breadth of the data is less extensive. The second one, by introducing popular machine learning techniques such as regression trees algorithms to examine whether these approaches provide superior forecasting results.

Chapter 1, entitled "Forecasting the equity risk premium in the European Monetary Union", investigates the capacity of multiple economic and technical variables to predict the Euro area equity risk premium. In the related literature, it is generally accepted that several economic and financial indicators show in-sample forecasting ability to predict stock returns. Nevertheless, there is not the same consensus when forecasting out-of-sample. Goyal and Welch (2003) evidenced that a long list of predictors from the literature could not perform consistently better out-of-sample than a naïve forecast based on the historical average. Later, other papers such as Campbell and Thompson (2008), Ferreira and Santa-Clara (2011) or Neely et al. (2014) showed that fixing some economically motivated restrictions or gathering the most relevant information in few principal components could beat the historical average's out-of-sample performance.

This chapter examines the performance of several variables that could be good predictors of the equity risk premium in the European Monetary Union for a period that spans from 2000 to 2020. In-sample, technical indicators display predictive power, matching or exceeding traditional economic forecasting variables. Nevertheless, out-of-sample exercises do not confirm in-sample results. Technical indicators do not show out-of-sample forecasting power, and only a few economic factors exhibit forecasting ability and provide economic value for a low risk-averse investor.

Chapter 2, entitled "Forecasting the European Monetary Union equity risk premium with regression trees", expands on the previous chapter. It investigates whether popular machine learning algorithms, such as classification and regression trees (CART), can help to improve

equity risk premium forecasts. For the moment, there is still not much literature exploring the ability of some of the most popular algorithms in data science to improve equity returns predictability. This second essay aims to contribute to this growing part of the literature focusing on a European dataset.

More precisely, the chapter investigates the capacity of three regression trees ensemble methods (bagging, random forests, and boosting) to select good economic predictors and to improve out-of-sample forecasts of the equity risk premium. As in the first chapter, the sample covers EMU monthly data from January 2000 to December 2020. Results obtained show that regression tree algorithms do not enhance forecasting ability and raise questions about machine learning algorithms' suitability when the data set dimension is not big enough. Moreover, it also explores whether these tree algorithms can provide economic value for a portfolio investor who chooses to invest in the risk-free asset or the equities market. Outcomes obtained are mixed, and not all performance indicators can confirm whether tree algorithms create economic value.

The third essay of this dissertation changes the financial asset to predict and investigates another prevalent theme in financial economics and banking. Since the Great Financial Crisis of 2007-2008, the underlying causes of non-performing loans and their relation with the economic cycles have become highly relevant for policymakers and practitioners in finance. The existing literature has repeatedly shown a clear inverse relationship between economic growth and defaults. In expansionary cycles, delinquency rates are usually lower or decrease, whereas, in recessions, they increase.

The bulk of this literature examines aggregated NPL data (adding up all types of credit portfolios), allowing to deal with large datasets with many cross-sectional observations. Nonetheless, little research exists at a more disaggregated level and analysing smaller dimension panel datasets, more in line with real situations that professionals sometimes have to address.

Chapter 3, entitled "Macro determinants of non-performing loans: A comparative panel analysis between consumer and mortgage loans", examines the influence of several macroeconomic factors on delinquency rates and contributes to the empirical literature in several ways. First, it adds research to the "large T, small N" panel literature on the economic determinants of non-performing loans. Second, it examines whether macroeconomic factors impact differently across different loan categories. Third, it contributes to the debate of whether

to pool or not to pool the data when the panel is unbalanced, heterogeneous, and not all the cross-sections are long enough to undertake time series analysis. Fourth, it wonders about the possibility of using fixed-effects models in dynamic panels with a long number of time observations.

Chapter 3 includes a dataset with quarterly data from 1992 to 2019 for a sample of 8 developed and emerging economies: United States, Spain, Mexico, Turkey, Colombia, Peru, Argentina and Chile. We selected this heterogeneous and short sample of countries as a possible real example of geographies and datasets analysed by practitioners in the financial industry. Results obtained favour pooled estimations, provide strong evidence supporting macroeconomic factors as crucial drivers of delinquency rates, and find similar estimation results using either static panel estimation techniques or dynamic ones.

Finally, we present an overview of the main contributions and results of this dissertation.

Chapter 1

Can we forecast the equity risk premium in the European Monetary Union?

1.1 Introduction

Forecasting stock returns is one of the most popular themes for both academics and practitioners in finance. The existing literature has studied many types of variables and proposed multiple econometric models to see whether there is significant evidence of returns predictability. It is generally accepted among financial economists that stock returns contain a significant predictable component in-sample. For example, Rozeff (1984), Campbell and Shiller (1988a) or Cochrane (2008) find evidence in favour of return predictability using the dividend yield. Campbell and Shiller (1988b,1998) use the earnings-price ratio; Fama and Schwert (1977) and Ang and Bekaert (2007) use nominal interest rates; Campbell and Vuolteenaho (2004) use inflation; Guo (2006) uses stock volatility; and many other studies use different economic variables that support stock returns predictability in-sample.

Nonetheless, when we review economic literature that focuses on out-of-sample forecasting power, economists have no consensus. Bossaerts and Hillion (1999), and Goyal and Welch (2003,2008), show that a long list of predictors from the literature cannot perform consistently better out-of-sample than a simple forecast based on the historical average. These studies conclude that the evidence found at the in-sample level does not hold up out-of-sample. On the other hand, other recent studies provide some evidence favouring stock returns predictability out-of-sample. Campbell and Thompson (2008) and Ferreira and Santa-Clara (2011) show that traditional literature predictors can beat the historical average's out-of-sample performance by fixing some economically motivated model restrictions. Neely et al. (2014) show out-of-sample forecasting ability using the economic variables studied by Goyal and Welch (2008) and several technical indicators.

We contribute to the literature by analysing the predictability of European stock returns. Despite the numerous literature focusing on the predictive ability of aggregate economic indicators in the US stock market, there have been fewer attempts to examine the predictive power in the European market as a whole. This gap is not surprising due to the lack of historical data if we compare it with the US market. Our analysis using European Monetary Union (EMU) data sheds new light on the nature of stock return predictability.

This paper investigates the capacity of multiple economic variables and technical indicators to forecast the equity risk premium. In particular, we work with monthly European Monetary Union data for a period that spans from the EMU foundation in 2000 to 2020. We

selected our methodology because we think it similarly approaches the stock return predictability as practitioners in finance do. It works with a wide range of predictors that try to gather the economy's state, produce out-of-sample forecasts, and use them as inputs for asset allocation decisions.

Following the research line started by Neely et al. (2014), the article first analyses the monthly in-sample forecasting ability of economic and technical variables, starting with traditional simple linear regressions. Simple linear regressions may not be enough to model equity risk premiums accurately because other predictors could incorporate relevant information. Hence we also estimate multivariate predictive regressions. However, instead of using all the variables, we use principal components analysis for summarising our predictors' relevant information into a few principal components.

Once the in-sample forecasting exercise is finalised, we continue with an out-of-sample one to check whether in-sample results were consistent. The whole sample is divided into two subsamples. The first one, the estimation period, expands from January 2000 to December 2012. The second, the out-of-sample forecasting window from January 2013 to December 2020. Forecasts are calculated monthly, and once a one-period equity risk premium is forecasted, the period's data is added to the estimation window, making the estimation window grow period after period. Predictive regressions are compared to a popular benchmark in the literature: the historical mean average. And forecasts performance is analysed in terms of the Campbell and Thompson R^2 (R^2_{os}), which compares the MSFE of regressions constructed with selected predictors against the MSFE of the benchmark.

To conclude, the economic value of the out-of-sample predictions are measured for a risk-averse investor with a quadratic utility function. We maximise investors' utility function using a simple asset allocation model that invests in equities or a risk-free asset. If predictive regressions improve investors' utility relative to the benchmark predictions, then it will be stated that those predictors create economic value for this investor.

Our results can be of interest because they support some of the previous findings achieved by Neely et al. (2014) for the US equity market. In particular, in the EMU, almost all technical predictors and principal components multivariate regressions exhibit statistically significant in-sample power. However, contrary to them, our out-of-sample results do not confirm in-sample ones. Technicals do not predict better than the benchmark out-of-sample. Only a few economic

and technical can produce economic value for low risk-averse investors, and almost none of them when investors show higher degrees of risk aversion.

The remainder of this chapter is organised as follows. Section 1.2 studies the forecasting power of the explanatory variables in-sample. Section 1.3 focuses on the out-of-sample analysis, explaining the methodology, and later reporting empirical results. Section 1.4 addresses a final exercise to measure the economic value of the out-of-sample forecasts for risk-averse investors. Section 1.5 puts together and summarises all the results obtained in previous parts. Finally, section 1.6 concludes.

1.2 In-sample analysis

1.2.1 Simple linear regressions

The conventional framework for analysing the equity risk premium predictability is based on simple linear regressions. The idea is to run a predictive linear regression of realised excess returns on lagged explanatory variables.

$$r_{t+1} = \alpha_i + \beta_i X_t + \varepsilon_{t+1} \quad (1)$$

Where the equity risk premium, r_{t+1} is the return on a broad stock market index above a risk-free asset from period t to t+1, x_t is the predictor, and ε_{t+1} is a zero-mean disturbance.

The challenge in this simple model is to select which variables to include on the right side of the equation since results can change substantially depending on which ones are used to take the role of explanatory variables.

The literature compiles evidence on numerous economic variables that have the ability to forecast the equity risk premium. One of the most popular predictors in this literature is the dividend price ratio. Papers such as Rozeff, (1984), Cambell and Shiller (1998a,1998), Fama and French (1988), Cochrane (2008), or Stambaugh (2009) work with dividends price ratios. Other variables often used by the literature are the earnings-price ratio (Campbell and Shiller (1998b,1998)), book-to-market ratio (Kothari and Shanken (1997)), nominal interest rates (Ang and Bakeert (2007)), interest rate spreads (Campbell (1987)), inflation (Campbell and Vuolteenaho (2004)), and several other variables detailed in Appendix 1.

In this article, we follow Neely et al. (2014) paper and analyze the European Monetary Union (EMU) to see if we can get similar results as they obtained for the United States. We work with 12 economic variables[1] and 14 technical indicators.

Data spans from January 2000 to December 2020, and frequency is monthly. There are two principal reasons to choose this data window. The first one is data availability. The EMU started in 1999, and many of the economic variables at the EMU level begin at this moment. The second is the existence of potential structural problems in the data. The monetary union represents a significant structural change in Europe, and data after this event might be affected by this structural break.

The equity risk premium is calculated as the difference between the monthly natural logarithm of the equity return and the risk-free asset. We selected a popular stock market index, the MSCI EMU index, to estimate the equity return. And as a risk-free asset, we chose the Euribor 1M, a European interbank index rate with a monthly maturity.

The set of economic variables used to forecast the equity risk premium are:

1. *Dividend-Price ratio (log), DP:* The ratio between the log of the past 12 month *I/B/E/S*[2] dividend per share for the MSCI EMU and the log of the *I/B/E/S* MSCI EMU index.

2. *Dividend-Yield ratio (log), DY:* The ratio between the log of the past 12 month *I/B/E/S* dividend per share for the MSCI EMU and the log of the *I/B/E/S* MSCI EMU index lagged one period.

3. *Earnings-Price ratio (log), EP:* The ratio between the log of the past 12 month *I/B/E/S* earnings per share for the *I/B/E/S* MSCI EMU and the log of the *I/B/E/S* MSCI EMU index.

4. *Payout ratio (log), DE:* The ratio between the log of the past 12 month *I/B/E/S* dividend per share for the MSCI EMU and the log of the past 12 month *I/B/E/S* earnings per share for the MSCI EMU.

5. *Book-to-Market ratio, BM: I/B/E/S* book to market ratio for MSCI EMU.

[1] Neely et al. (2014) works with 14 economic variables, but for the EMU area we were not able to replicate 2 of them: Net equity expansion and Long-term return.

[2] Institutional Brokers' Estimate System (*I/B/E/S*) is a service that gathers and compiles stock data and analyst estimates.

6. *Euribor 3M rate, EUR3M.*

7. *Swap 10Y, SWAP10:* 10 years Euro Swap rate.

8. *Term Spread, TMS:* It is the difference between the 10 years swap rate and the Euribor 3M rate.

9. *Default Yield Spread, DFY:* The difference between the asset swap spread of EMU High Yield index (Bank of America Merrill Lynch EMU Corporate Index) and the spread of EMU Investment Grade index (Bank of America Merrill Lynch Euro High Yield index).

10. *Default Return Spread, DFR:* The difference between the yield to maturity of EMU High Yield index (Bank of America Merrill Lynch EMU Corporate Index) and the yield to maturity of EMU Investment Grade index (Bank of America Merrill Lynch Euro High Yield index).

11. *Inflation, INF:* EU Harmonized CPI Y/Y change (NSA)

12. *Equity Risk Premium volatility, RVOL:* It can be defined as:

$$Vol_t = \sqrt{\frac{\pi}{2}}\sqrt{12\sigma_t}$$

where

$$\sigma_t = \frac{1}{12}\sum_{i=1}^{12}|r_{t+1-i}|$$

Table 1 reports summary descriptive statistics for the equity risk premium and the economic variables. Surprisingly, the selected period's equity risk premium is negative, which contradicts most past literature. Our explanation for this puzzle is simple. The chosen period (Jan 2000 – Dec 2020) is short, and the last financial crisis and Covid-19 correction weigh heavily in the data.

Table 1: Summary statistics

	Mean	Stdev	Min	Max	Autocorr	Sharpe
Equity Premium (%)	-0,15	5,26	-19,55	15,79	0,09	-0,03
DP	0,35	0,06	0,20	0,46	0,99	
DY	0,35	0,06	0,20	0,47	0,99	
EP	0,47	0,06	0,34	0,58	0,99	
DE	0,73	0,06	0,56	0,83	0,98	
RVOL	0,17	0,07	0,07	0,38	0,97	
BM	0,63	0,15	0,32	1,08	0,97	
EUR3M	1,58	1,79	-0,55	5,28	1,00	
SWAP10A	2,81	1,81	-0,28	5,95	1,00	
TMS	1,23	0,71	-0,57	2,88	0,96	
DFY	3,98	2,00	1,36	12,05	0,96	
DFR	4,86	3,39	1,40	17,68	0,97	
INF	1,65	0,94	-0,62	4,08	0,97	

The sharpe ratio is calculated as the ratio of the mean return and the standard deviation.

Technical variables used to forecast the equity risk premium are derived from 3 types of technical strategies:

1. *Moving averages rules*. These rules give buy and sell signals depending on the short and long moving averages of prices. If the short moving average is above the long average, then there is a buy signal (S_t=1), and if it is below, there is a sell signal (S_t=0). We analyse monthly MA rules with short MA with t= 1,2, 3 months, and long MA with t=9,12. This gives six technical predictors: MA1MA9, MA1MA12, MA2MA9, MA2MA12, MA3MA9 and MA3MA12.

$$S_t = \begin{cases} 1 \; if \; MA_{short} \geq MA_{long} \\ 0 \; if \; MA_{short} < MA_{long} \end{cases}$$

where

$$MA_t = \frac{1}{t}\sum_{i=1}^{t} p_i$$

and

$$p_i = level \; of \; stock \; price \; index$$

Moving average rules help to find trends and recognise changes in those trends. When short moving averages stay above the long ones, stock prices show an uptrend. On the other hand, if short moving averages are lower than long ones, stock prices offer a downtrend. Points at which short averages cross long averages identify possible trend changes.

2. *Momentum rules*. If the current stock price is higher than its level m periods before, that gives a positive momentum (S_t=1). If the current price is lower, then we have a negative momentum (S_t=0). We compute monthly signals for m=9,12, which gives two technical predictors: MoM9 and MoM12.

$$S_t = \begin{cases} 1 \; if \; p_t \geq p_{t-m} \\ 0 \; if \; p_t < p_{t-m} \end{cases}$$

3. *Volume Rules*. The paper works with the "on balance" volume (OBV) indicator that was developed by Granville (1963). It is a momentum indicator that relates volume to stock price change. It measures buying and selling pressure as a cumulative indicator that adds volume on up days and subtracts volume on down days.

$$\text{If } p_t \geq p_{t-1} \text{ then } OBV_t = Volume_t$$
$$\text{If } p_t < p_{t-1} \text{ then } OBV_t = -Volume_t$$

The OBV is calculated as:

$$OBV = \text{Cumulative up to } OBV_{t-1} + OBV_t$$

Then we form a trading signal using moving averages of the OBV. If the short moving average is above the long average, then there is a buy signal (S_t=1), and if it is below, there is a sell signal (S_t=0). We analyse monthly MA rules with short MA with t= 1,2, 3 months, and long MA with t=9,12. This gives six technical predictors: OBV19, OBV112, OBV29, OBV212, OBV39 and OBV312.

$$S_t = \begin{cases} 1 \; if \; MA_{short}^{OBV} \geq MA_{long}^{OBV} \\ 0 \; if \; MA_{short}^{OBV} < MA_{long}^{OBV} \end{cases}$$

where

$$MA_t = \frac{1}{t} \sum_{i=1}^{t} OBV_i$$

The idea behind the OBV measure is that volume movement is thought to happen before price movements. Therefore, situations where volumes are falling and prices are still rising are selling pressure signals. And the opposite is true when volumes are up, and prices are down.

Most of the economic literature forecasting the equity risk premiums does not include technical indicators. However, these indicators can capture relevant information which macroeconomic and fundamental variables cannot gather. For example, Treynor and Ferguson (1985) show technical indicators can help assess whether all information has been incorporated into equity prices. Moreover, technical indicators can better capture price trends than other variables. Cespa and Vives (2012) show that asset prices can deviate from their fundamental values for extended periods if there is a positive level of asset residual payoff uncertainty or persistence in liquidity trading. In summary, there are important reasons that can help to explain why technical indicators display some predictive ability.

1.2.2 Multivariate predictive regressions

Simple linear regressions are often not sufficient to model accurately dependent variables. In this sense, multivariate regressions could improve the model's accuracy by adding relevant information that new variables could provide. However, using all the above predictors in a regression could create more problems than benefits (e.g. multicollinearity, loss of degrees of freedom, in-sample overfitting, and so on). Thus, to reduce the number of correlated economic and technical indicators while preserving most of the information of the large set of variables, principal components analysis (PCA) is applied.

PCA is a technique that searches for few uncorrelated linear combinations of the original variables and captures most of those variables' information. The principal components are ordered concerning their variation so that the first few components, if actual variables, show strong linear relation, account for most of the variation. Or, equivalently, the few first principal components identify the key comovements among the entire set of predictors, which filters out much of these ones' noise.

Considering these fundamental characteristics of PCA, we have calculated three sets of principal components. One for the economic variables, F^{Econ}, other for the technical variables, F^{Tech}, and a final group that includes all the used variables, F^{All}.

Principal components predictive regressions are given by

$$r_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k F_{k,t}^{J} + \varepsilon_{t+1} \quad (2)$$

Where $F_{k,t}^{J}$ is the k principal component at time t, and J can be equal to Econ, Tech or All, depending on the subset of variables we are working with.

One of the main difficulties in using PCA is selecting the number of components. There are plenty of methods to calculate the number of principal components, but to keep things more straightforward, we use a rule of thumb: to select components that explain 80% of the data variance.

Figures 1 and 2 visualise the PCA results applied to the economic, technical, and all variables. Figure 1 plots the linear combination of the variables in each of the components selected, whilst Figure 2 represents the loadings for each component.

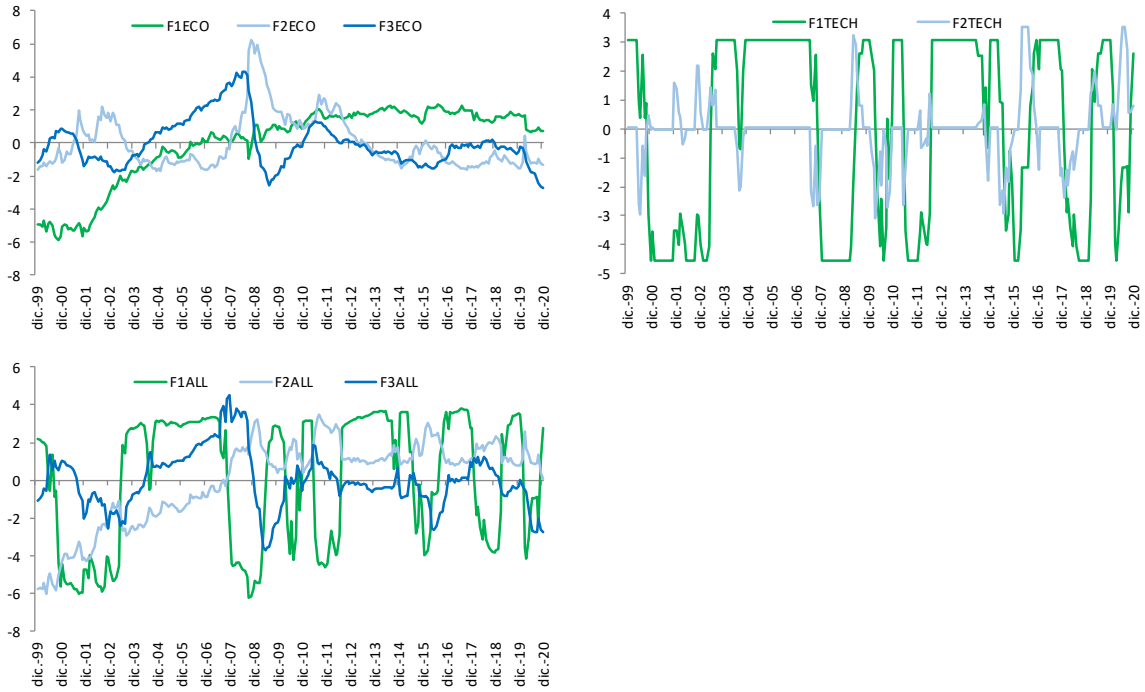


Figure 1: Representation of variables in the principal components space

Figure 2 illustrates which variables dominate in each extracted component. For economic variables, valuation indicators and interest rates dominate the first component PC1 ECON, which explains around 45% of the common variance. The second component, PC2 ECON, adds another 22% of the variance, and it is better explained by risk measures such as the

volatility of equity prices or default risk measures (DFY and DFR). For technical variables, the first component PC1 TECH explains almost 80% of the common variance, and all the variables contribute at a similar level to the component. There is no clear dominance of any of them, which points to a strong linear correlation among all the technical variables. Finally, taking all predictors together, Figure 2 shows that technical variables load heavily in the first component PC1 ALL, while valuation and interest rates load further in the second component PC2 ALL.

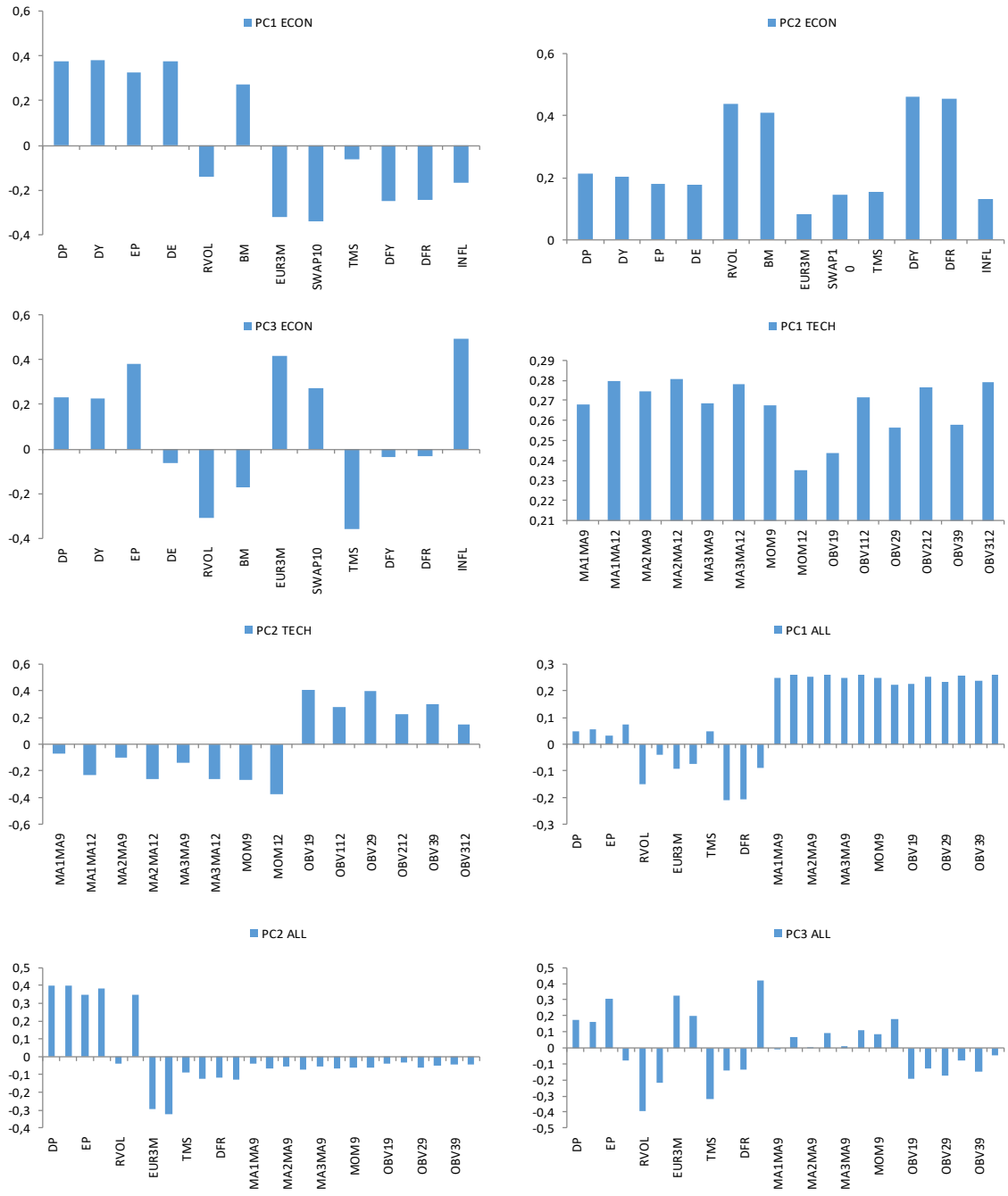


Figure 2: Loadings on principal components, January 2000 to December 2020

1.2.3 In-sample estimations and equity risk premium forecasting ability

To estimate predictive regressions, we use simple OLS estimators. However, using OLS estimators with predictive regressions can give several statistical concerns. The first problem is that regression residuals obtained through OLS estimation may show serial correlation and heteroskedasticity, making OLS estimators biased and inconsistent. A second problem is the well-known Stambaugh bias (1986, 1999). This widespread bias in predictive regressions arises when the predictor is highly persistent, and the predictor and return innovations are correlated. Both problems make conventional t-tests for testing the hypothesis of no predictability $\beta=0$ unusable, given that standard errors are not correctly calculated and traditional distributions are no longer valid for hypothesis testing or construction of confidence intervals.

The first problem, serially correlated errors and heteroskedasticity is sorted out using Newey and West (1987) standard errors. The Stambaugh bias is addressed with the solution proposed by Neely et al. (2014). They solved the Stambaugh bias by computing p-values using a wild bootstrap procedure that accounts for the persistence in regressors and correlations between equity risk premium and predictor innovations, as well as general forms of heteroskedasticity.

We run predictive regressions (simple linear regressions for individual predictors and multivariate linear regressions for principal components), obtain all the estimators, and test for H_0: $\beta_i=0$ versus H_A: $\beta_i\neq0$. Table 2 displays these regressions and hypothesis tests from January 2000 to December 2020.

Table 2: In-sample estimation results

Economic Variables					Technical Variables					Principal Components Variables					
Predictor	**Beta**	**t-stat**	**Prob**	**R2 (%)**	**Predictor**	**Beta**	**t-stat**	**Prob**	**R2 (%)**	**Predictor**	**Beta**	**t-stat**	**Prob**	**R2 (%)**	**F-Stat**
DP	1,4	0.2	0,85	0,0	**MA1MA9**	**1,5**	**2.0****	**0,05**	**1,9**	**F1ECO**	**0,3**	**1.8***	**0,05**	**5,8**	**5.1****
DY	2,2	0.3	0,75	0,1	**MA1MA12**	**1,4**	**1.9****	**0,07**	**1,7**	**F2ECO**	-0,1	-0.6	0,62		
EP	0,8	0.1	0,94	0,0	**MA2MA9**	**1,4**	**1.9****	**0,08**	**1,7**	**F3ECO**	**-0,8**	**-2.8*****	**0,01**		
DE	3,9	0.7	0,48	0,2	**MA2MA12**	0,8	1.0	0,30	0,5	**F1TECH**	**0,2**	**2.1****	**0,04**	**2,5**	**3.1****
RVOL	-0,2	0.	0,98	0,0	**MA3MA9**	**1,5**	**2.0****	**0,05**	**1,9**	**F2TECH**	0,3	1.1	0,27		
BM	**6,0**	**2.6*****	**0,01**	**2,9**	**MA3MA12**	0,8	1.0	0,31	0,5	**F1ALL**	**0,2**	**2.2****	**0,02**	**6,5**	**5.7*****
EUR3M	**-0,6**	**-3.1*****	**0,00**	**4,3**	**MOM9**	**1,4**	**1.9****	**0,05**	**1,7**	**F2ALL**	0,2	1.2	0,27		
SWAP10Y	**-0,5**	**-2.4****	**0,02**	**2,5**	**MOM12**	0,4	0.6	0,55	0,2	**F3ALL**	**-0,7**	**-2.8****	**0,01**		
TMS	**0,9**	**1.7***	**0,10**	**1,4**	**OBV19**	0,8	1.1	0,28	0,6						
DFY	-0,3	-1.5	0,19	1,3	**OBV112**	**1,5**	**2.0****	**0,04**	**2,0**						
DFR	-0,2	-1.4	0,25	1,3	**OBV29**	**1,5**	**2.0****	**0,04**	**1,9**						
INF	**-1,2**	**-3.2*****	**0,00**	**4,5**	**OBV212**	**1,5**	**2.1****	**0,05**	**2,1**						
					OBV39	**1,7**	**2.5****	**0,01**	**2,7**						
					OBV312	**2,1**	**2.9*****	**0,00**	**3,9**						

, ** and * indicate significance at the 10%, 5% and 1% levels, respectively, based on two-tail wild bootstrapped p-values.*

Table 2 shows that only five out of twelve predictors are significant within the economic variables and show in-sample predictive power. Among these five variables, four could be classified as macroeconomic factors, EUR3M, SWAP10, TMS and INF; and only one as a fundamental indicator, BM. EUR3M and INF exhibit the highest R^2. Given the predictive balance between the macroeconomic and fundamental factors, it could be stated that macroeconomic factors show higher in-sample predictive power in this sample. A possible explanation for this lack of forecasting power of fundamental predictors could be that these are better predictors for longer-term periods. As noted earlier, equity prices can deviate from fundamental values for extended periods (see Cespa and Vives (2012) or Hong and Stein (1999)), making these variables work better in more extended prediction periods.

Concerning technical indicators, many of them are significant and show forecasting ability. This result supports the idea that technicals may gather some investors' behaviour that economic variables cannot. It also suggests that some individual technical indicators can predict the equity risk premium as good as individual economic variables. If we look at their R^2, volume indicators such as OBV312, OBV39 and OBV212 display the highest forecasting power with R^2s of 3.9%, 2.7%, and 2.1%. Though this R^2 may seem small, Campbell and Thomson (2008) and Neely et al. (2014) claim that a monthly R^2 near 0.5% can represent an economically significant degree of equity risk premium predictability.

Multivariate outcomes also show that all three regressions are significant. Performance results are better for the regressions constructed with principal components than using individual predictors, and the best performing regression is the one that combines in three principal components all the 26 individual predictors, with an R^2 of 6.5%. Hence, combining the different information provided by economic and technical variables could improve predictive power.

Finally, if we compare our results to the results obtained by Neely et al. (2014) for the US, these are similar. First, both results found that plenty of technical predictors exhibit statistically significant in-sample predictive power. Second, combining information from both technical and economic indicators using PCA produces superior in-sample forecasts. Moreover, for the economic variables, both studies found that three of the most studied predictors in the literature, short term yields (Euribor3M), long term yields (Swap10Y) and the yield curve (Swap10Y-Euribor3M), are statistically significant. Nonetheless, our results do not match for value

indicators. Neely et al. (2014) found that dividend yields have significant forecasting power in-sample, and we could not find the same result in th EMU.

1.3 Out-of-sample analysis

Following Neely et al. (2014) and as a robustness check, we run out-of-sample forecasting tests to check if in-sample results are also valid out-of-sample.

To calculate out-of-sample forecasts, we divide the whole data period into two subperiods. The first one, known as the estimation period, spans from January 2000 to December 2012. The other one, the out-of-sample subperiod, expands from January 2013 to December 2020. The initial estimation window is selected as long as possible to get precise and stable estimators, and this initial window grows as new data is added[3].

The out-of-sample equity premium forecast based on individual variables is given by:

$$\hat{r}_{t+1} = \hat{\alpha}_{t,i} + \hat{\beta}_{t,i} X_{i,\,t} \tag{3}$$

Where $\hat{\alpha}_{t,i}$ and $\hat{\beta}_{t,i}$ are the estimates from regressions for economic or technical variable i, up to month t.

The out-of-sample forecast based on principal components is given by:

$$\hat{r}^{J}_{t+1} = \hat{\alpha}_t + \sum_{k=1}^{K} \hat{\beta}_{t,k} F^{J}_{t,k} \quad \text{for J= ECON, TECH or ALL} \tag{4}$$

Where $F^{J}_{t,k}$ is the *k*th principal component extracted from the economic, technical, or all variables.

The accuracy of the proposed predictive regressions forecasts is compared to a benchmark to analyse out-of-sample stock return predictability. A very popular benchmark used in the literature (e.g. Goyal and Welch (2008), Campbell and Thomsom (2008) or Ferreira and Santa-Clara 2011) is the historical mean average.

[3] First we choose the initial estimation window and calculate the out-of sample forecasted excess return and compare it to the real excess return. For the next period, the estimation window adds the new data and we repeat the whole process. We continue doing this until the end of the data set is reached and we have generated (T-τ) out-of-sample forecasts, where T is the whole sample, and τ is the out-of-sample subperiod.

$$\hat{r}_{t+1}^{HA} = \frac{1}{t}\sum_{i=1}^{t} r_i \quad (5)$$

This benchmark forecast assumes that the best predictor for the equity risk premium at t+1 is the risk premium's historical average up to period t.

A prevalent metric is the Mean Squared Forecast Error (MSFE) to evaluate forecast accuracy. In line with Neely et al. (2014), we analyse forecasts in terms of the Campbell and Thomson (2008) out-of-sample R^2 (R_{os}^2), which compares the MSFE of regressions constructed with selected predictors to the MSFE of the benchmark, and it can also be compared with the in-sample R^2 statistic (see Campbell and Thompson (2008)).

$$R_{os}^2 = 1 - \frac{MSFE_r}{MSFE_b} \quad (6)$$

Where $MSFE_r$ is the mean squared forecast error of the predictive regression, whilst $MSFE_b$ is the mean squared forecast error of the benchmark. This R_{os}^2 measures the proportional reduction in the MSFE for the predictive regression forecast relative to historical averages. Thus, if $R_{os}^2 > 0$, then the predictive regression forecast relative to the historical average is more accurate. But If $R_{os}^2 < 0$, then the opposite happens, and the predictive regression forecast cannot beat the benchmark.

We are also interested in determining whether the improvement in the forecast is also significant, that is, testing the null hypothesis H_0: $MSFE_b \leq MSFE_r$ versus the alternative hypothesis H_a: $MSFE_b > MSFE_r$, which is the same as testing H_0: $R_{os}^2 \leq 0$ versus H_a: $R_{os}^2 > 0$.

Diebold and Mariano (1995) and West (1996) proposed a statistic (DMW statistic) for testing the null of equal MSFE between the two models. They proved that when comparing forecasts from non-nested models, DMW has a standard normal asymptotic distribution, and it can be compared to typical critical values of 1.282, 1.645 and 2.326 for the 10%, 5% and 1% significance levels, respectively.

However, in the context of our predictive regressions, we are working with nested models since our predictive regressions (equations 1 and 2) reduce to the constant expected excess return model when $\beta_i=0$ (equation 5). And as Clark and McCracken (2001) proved, DMW statistic has a non-standard asymptotic distribution when comparing forecasts from nested models. Hence, comparing the DMW statistic against standard normal critical values usually makes tests undersized.

Clark and West (2007) found a solution for nested models and proposed an MSFE-Adjusted statistic[4] that provides a method for assessing statistical significance. These Clark and West (2007) MSFE-Adjusted statistics, R^2_{os}, and Theil[5] decomposition of MSFE are reported in Table 3.

Table 3: Out-of-sample forecasting results

		MSFE % Decomposition		
Predictor	**MSFE**	**Bias**	**Variance**	**Covariance**
HA	21,3	2,4%	92,9%	4,7%

					MSFE % Decomposition		
Predictor	**MSFE**	**R2OS**	**t(MSFE)**	**P-Value**	**Bias**	**Variance**	**Covariance**
DP	21,3	0,1	0.72	0,23	2,1%	92,8%	5,1%
DY	21,3	0,3	1.05	0,15	2,0%	93,2%	4,8%
EP	21,3	0,1	0.34	0,37	2,2%	92,9%	4,9%
DE	21,2	0,6	1.26	0,10	1,6%	91,6%	6,8%
RVOL	21,4	-0,3	-0.61	0,73	2,0%	89,3%	8,7%
BM	**20,4**	**4,5**	**3.44*****	0,00	**1,5%**	**90,9%**	**7,7%**
EUR3M	**22,1**	**-3,7**	**1.71***	0,04	**5,6%**	**80,7%**	**13,7%**
SWAP10Y	**22,5**	**-5,5**	**1.67***	0,05	**6,6%**	**72,9%**	**20,5%**
TMS	22,3	-4,5	-0.81	0,79	5,3%	75,3%	19,4%
DFY	21,2	0,6	1.12	0,13	0,1%	86,8%	13,1%
DFR	21,1	1,1	1.27	0,10	0,2%	90,0%	9,8%
INF	**21,7**	**-1,7**	**1.74***	0,04	**2,0%**	**55,0%**	**43,0%**

					MSFE % Decomposition		
Predictor	**MSFE**	**R2OS**	**t(MSFE)**	**P-Value**	**Bias**	**Variance**	**Covariance**
FECO	**21,9**	**-2,6**	**1.88****	0,03	**5,2%**	**70,9%**	**24,0%**
FTECH	22,8	-7,0	-0.13	0,55	0,6%	52,2%	47,2%
FALL	**21,9**	**-2,8**	**1.43***	0,08	**0,2%**	**54,5%**	**45,2%**

					MSFE % Decomposition		
Predictor	**MSFE**	**R2OS**	**t(MSFE)**	**P-Value**	**Bias**	**Variance**	**Covariance**
MA1MA9	22,6	-6,1	-0.28	0,61	1,5%	56,0%	42,5%
MA1MA12	22,3	-4,8	-0.03	0,51	1,4%	57,4%	41,2%
MA2MA9	22,1	-3,8	0.06	0,47	1,5%	60,0%	38,5%
MA2MA12	22,5	-5,7	-0.94	0,83	1,3%	65,1%	33,6%
MA3MA9	22,0	-3,3	0.21	0,42	1,7%	59,5%	38,9%
MA3MA12	22,0	-3,3	-0.34	0,63	1,6%	68,1%	30,3%
MOM9	22,1	-3,8	0.25	0,40	1,6%	56,8%	41,6%
MOM12	21,8	-2,4	-0.41	0,66	1,6%	74,2%	24,1%
OBV19	22,4	-5,1	-0.83	0,80	1,1%	66,9%	32,1%
OBV112	22,7	-6,7	-0.27	0,61	0,6%	55,6%	43,9%
OBV29	22,5	-5,7	0.00	0,50	0,9%	54,6%	44,5%
OBV212	22,5	-5,5	0.16	0,44	0,7%	52,9%	46,4%
OBV39	22,0	-3,1	0.58	0,28	0,7%	55,5%	43,8%
OBV312	21,5	-0,7	1.27	0,10	0,6%	52,7%	46,7%

, ** and * indicate significance at the 10%, 5% and 1% levels, respectively.*

Out-of-sample outcomes contradict some of the conclusions obtained in-sample. The most stinking result is that none of the technical indicators shows forecasting ability out-of-sample. The Campbell and Thompson R^2_{os} are negative for every variable, and Clark and West tests indicate that none of the predictors shows an MSFE significantly lower than the benchmark. These results strongly contradict what we found in-sample, where almost all technical predictors showed forecasting ability. Besides, they disagree with Neely et al. (2014), who find that individual technical indicators perform out-of-sample as well or better than economic variables.

[4] Clark and West (2007) propose, instead of comparing MSFEb to MSFEr for the hypothesis testing, to compare MSFEb to an adjusted MSFEr that corrects for the noise associated with the larger model forecast. Comparing to this adjusted MSFEr would allow to use normal critical values (1.282, 1.645, and 2.326).

[5] The Theil (1971) MSFE decomposition is given by: a) the bias component $\left(\hat{\bar{X}} - \bar{X}\right)$, that represents how far is the mean of the forecast from the mean of the actual series; b)the variance component $(S_{\hat{X}} - S_X)$, which measures how far the variance of the forecast is from the actual series; and c) the covariance component $2(1 - r_{\hat{X}X})S_{\hat{X}}S_X$, where r is the correlation, and this component measures the remaining unsystematic forecasting error. If the forecast is good, the bias and the variance proportions should be small so that most of the bias should be concentrated on the covariance proportion.

Out-of-sample performance of the economic variables is more in line with results obtained in-sample. Four (BM, EUR3M, SWAP10Y, INF) of the five variables which had in-sample power to predict have the out-of-sample too. Indeed, three of these predictors exhibit negative R_{os}^2 but MSFE-Adj statistics indicate that the MSFEs are significantly less than that of the historical average. This result is entirely possible when comparing nested model forecasts, and the MSFE-adj statistic can reject the null hypothesis even if the R_{os}^2 is negative. Similarly, six indicators (DP, DY, EP, DE, DFY and DFR) have positive R_{os}^2 but we cannot reject the hypothesis testing that benchmark forecast errors are lower than predictive regressions. Therefore we can not statistically confirm that these six predictors exhibit out-of-sample explanatory power.

Another angle to understand the out-of-sample ability to forecast can be found by looking at the Theil (1971) MSE decomposition. Theory suggests that forecasts are acceptable if parts explained by the bias and the variances proportions are small, and most of the MSE is defined by the unsystematic component proportion. In general, the MSFE decomposition points to a weak forecasting ability for most economic predictors. They exhibit the ability to forecast the series's mean but perform poorly when forecasting standard deviations. Nonetheless, when compared to the benchmark, historical averages show more flawed forecasting power than forecasting regressions. The average forecast's bias proportion is 2.4%, whilst the variance is 92,9%. BM and INF display lower bias and variance proportions, whilst EUR3M and SWAP10Y show bias proportions above the benchmark but lower variances. Technical indicators show a better MSFE decomposition, with more downward bias and variance ratios and higher covariance. Nonetheless, R_{os}^2 and Clark and West (2007) test rejected any forecasting power.

To conclude, the achievements of the principal components are mixed too. Results suggest that gathering common information from economic and all (economic and technical) variables improves forecasting ability out-of-sample. MSFE-adjusted statistics are significant for FECO and FALL, and Theil decomposition suggests a much better forecasting power than the benchmark. In particular, FALL decomposition shows a bias and a variance proportion well below the benchmark. However, contrary to what was found in-sample, technical variables FTECH, do not exhibit out-of-sample forecasting ability.

1.4 Asset allocation

As a final exercise, following Neely et al. (2014) and Ferreira and Santa-Clara (2011), we measure the economic value for a risk-averse investor, with a quadratic utility function, who invests her wealth combining equity with the risk-free asset.

The investor tries to maximise her expected quadratic utility function. In terms of a risky portfolio P, it can be defined as:

$$Max\ [E(U)] = Max\ E\left[\hat{\mu}_p - \frac{\delta}{2}\hat{\sigma}_p^2\right] \tag{7}$$

Where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the mean and the variance, respectively, for the investor's portfolio over the forecast evaluation period, and δ measures her degree of risk aversion.

The utility function can also be interpreted as an approximation of the certainty equivalent return (CER). CER would be the minimum guaranteed return that a mean-variance investor with a risk-aversion coefficient δ would consider equivalent to investing in the strategy. Put it more straightforwardly; the CER is the guaranteed rate an investor is willing to accept rather than taking a particular trading strategy. A higher CER means that a specific risk-averse investor needs higher compensation not to take the strategy.

To calculate its expected utility and maximise it, the investor first calculates each period her expected excess returns and finds the Markowitz optimal weights on the stock market through the following equation:

$$w_t = \left(\frac{1}{\delta}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}}\right) \tag{8}$$

Where $\hat{r}_{t+1}$ is the forecasted equity risk premium for the next period given by each strategy, and $\hat{\sigma}_{t+1}$ is the variance of the equity risk premium that we estimate using available past data. In line with Neely et al. (2014) and Campbell and Thompson (2008), we impose three assumptions: First, we assume that the investor uses a five-year rolling window to calculate the equity risk premium variance. Second, we constrain w_t to lie between 0 and 1.5,

limiting short sales and allowing a maximum of 50% leverage. Third, we forecast the equity risk premium in the same way we did in section 1.3 to calculate out-of-sample performance[6].

Once the investor knows her optimal weights between the risky asset (w_t) and the risk-free asset ($1 - w_t$), then she can calculate portfolio returns at the end of each period as:

$$r_{p,t+1} = w_t r_{t+1} + r_f \tag{9}$$

Where r_f denotes the risk-free return from time t to t+1, r_{t+1} is the equity risk premium for the period t+1, w_t is the given weight by each strategy to the equity index at period t, and $r_{p,t+1}$ is the portfolio return at period t+1.

Whether trading strategies have economic value or not will be determined by the investor's CER. If the CER obtained using predictive regressions is superior to the CER obtained from the benchmark, then there is economic value in the predictive regression and vice-versa. Hence, to evaluate the performance of each strategy, we calculate its CER:

$$\widehat{CER} = \hat{\mu}_p - \frac{\delta}{2}\hat{\sigma}_p^2 \tag{10}$$

Where $\hat{\mu}_p$ represents the sample mean of the estimated portfolio excess returns for each strategy, $\hat{\sigma}_p^2$ is the sample variance of those estimated excess returns, and δ measures the investor's coefficient of relative risk aversion.

Along with the CER, we use other two criteria to compare the out-of-sample performance of the different trading strategies:

1) The out-of-sample Sharpe ratio, defined as the sample mean of the out-of-sample excess returns, $\hat{\mu}_p$, divided by their sample standard deviation, $\hat{\sigma}_p$.

$$\widehat{SR} = \frac{\hat{\mu}_p}{\hat{\sigma}_p} \tag{11}$$

The out-of-sample portfolio turnover. It gives the amount of trading required to implement each portfolio strategy, and it is defined as the average sum of the absolute value of trades realised.

[6] First we choose the initial estimation window, Jan00 to Dec12, and calculate the out-of sample forecasted excess return for the next month. Then we add next period information and repeat the whole process. We continue doing this until the end of the data set is reached and we have generated (T-τ) out-of-sample forecasts, where T is the whole sample, and τ is the out-of-sample subperiod.

$$Turnover = \frac{1}{T-1}\sum_{t=\tau}^{T-1}\left(\left|w_{t+1}^{i} - w_{t+}^{i}\right|\right) \quad (12)$$

Where w_{t+}^{i} is the portfolio weight in the risky asset before rebalancing, but at t+1, w_{t+1}^{i} is the rebalanced portfolio weight in equities at t+1, and T is the total number of portfolio returns obtained in the estimation period.

Since we calculate the turnover for each trading strategy, it is possible to include trading costs into the portfolio returns, penalising those strategies with more trading operations. Formally, we compute portfolio returns, including transaction costs as:

$$r_{p,t+1} = [w_t r_{t+1} + r_f] - [TO_t x\ trading\ costs] \quad (13)$$

Where TO_t is the turnover of each strategy at time t, and trading costs represent each individual's transaction monetary cost.

Table 4 reports CER gain calculations, including proportional transactions costs equal to 50 basis points per transaction, out-of-sample Sharpe ratios, and out-of-sample portfolio turnovers, for each trading strategy and the historical average benchmark. It shows these three performance measures for three types of risk-averse investor: A highly risk-averse investor with a constant relative risk aversion (CRRA) coefficient of 10, an average risk-averse investor with a CRRA coefficient of 5, a low risk-averse investor with a CRRA coefficient of 1.

Table 4: Portfolio performance measures, January 2013 to December 2020

Predictor	CRRA=1			CRRA=5			CRRA=10		
	CER Gain (TC=50bps)	Sharpe	Turnover (%)	CER Gain (TC=50bps)	Sharpe	Turnover (%)	CER Gain (TC=50bps)	Sharpe	Turnover (%)
Benchmark									
HA	-0,2	0,00	0,0	-0,2	0,00	0,0	-0,2	0,00	0,0
Economic									
DP	-0,2	0,00	0,0	-0,2	0,00	0,0	-0,2	0,00	0,0
DY	-0,2	0,00	0,0	-0,2	0,00	0,0	-0,2	0,00	0,0
EP	-0,2	0,00	0,0	-0,2	0,00	0,0	-0,2	0,00	0,0
DE	-0,9	-0,16	0,2	-0,4	-0,16	0,0	-0,3	-0,16	0,0
RVOL	-0,4	-0,04	0,2	-0,3	-0,04	0,0	-0,2	-0,04	0,0
BM	**8,3**	**0,30**	**1,4**	**1,7**	**0,29**	**0,3**	**0,8**	**0,29**	**0,2**
EUR3M	**5,0**	**0,10**	**0,0**	-7,2	0,07	2,5	-5,8	0,05	4,0
SWAP10Y	**5,0**	**0,10**	**0,0**	-7,3	0,07	2,1	-7,1	0,05	4,2
TMS	**0,8**	**0,12**	**0,2**	0,0	0,12	0,0	-0,1	0,12	0,0
DFY	-3,0	-0,02	2,2	-1,8	-0,04	1,5	-1,0	-0,04	0,8
DFR	-3,7	-0,04	2,5	-1,5	-0,05	1,3	-0,9	-0,05	0,6
INF	**3,2**	**0,08**	**1,3**	-1,2	0,10	1,6	-3,3	0,09	2,7
Technical									
MA1MA9	**1,6**	**0,06**	**1,2**	-1,1	0,01	1,8	-0,6	0,01	0,9
MA1MA12	**2,1**	**0,07**	**1,3**	-0,9	0,01	1,7	-0,6	0,01	0,9
MA2MA9	**2,4**	**0,07**	**1,5**	-0,5	0,03	1,6	-0,4	0,03	0,8
MA2MA12	-3,9	-0,03	1,4	-2,0	-0,06	1,2	-1,1	-0,06	0,6
MA3MA9	**2,6**	**0,07**	**1,5**	-1,2	0,02	1,7	-0,7	0,02	0,9
MA3MA12	-2,1	-0,01	1,3	-1,4	-0,03	1,1	-0,8	-0,03	0,6
MoM9	**2,1**	**0,06**	**1,3**	-1,4	0,02	1,9	-0,8	0,02	0,9
MoM12	-4,0	-0,05	2,1	-1,2	-0,07	0,7	-0,7	-0,07	0,3
OBV19	-3,8	-0,03	2,5	-1,9	-0,06	1,3	-1,1	-0,06	0,7
OBV112	-0,3	0,03	1,3	-2,9	-0,02	2,5	-1,6	-0,02	1,2
OBV29	**0,7**	**0,04**	**1,1**	-2,3	0,00	2,3	-1,3	0,00	1,2
OBV212	**1,0**	**0,05**	**0,7**	-2,4	0,01	2,5	-1,3	0,01	1,2
OBV39	**3,2**	**0,08**	**1,1**	-1,7	0,03	2,6	-0,9	0,03	1,3
OBV312	**6,0**	**0,12**	**0,7**	-1,0	0,07	3,1	-0,6	0,07	1,5
Principal Components									
F ECO	**5,0**	**0,10**	**0,0**	-3,2	0,10	2,3	-3,9	0,09	3,7
F TECH	-0,3	0,02	1,2	-2,2	0,01	2,3	-1,2	0,01	1,1
ALL	**0,4**	**0,04**	**1,1**	-0,7	0,08	2,8	-0,6	0,08	2,0

CRRA is the Constant Relative Risk Aversion coefficient. It represents the investor's degree of risk aversion. Higher values mean higher risk aversion. CER Gain (TC=50bps) is the CER gain including transaction costs.
Transaction costs are calculated as TC x Turnover, where TC is the cost of a trade.
Sharpe = monthly sharpe ratio.
Turnover measures the % cost of rebalancing trades per month for each strategy.
Bold numbers denote a stategy CER higher than that of the benchmark, taking into account transaction costs.

Table 4 results indicate that few trading strategies offer a higher CER than the historical average, and these work better for the less risk-averse investor (CRRA=1) than for riskier ones (CRRA=5 or CRRA=10). For an investor with a risk-aversion coefficient equal to one, sixteen out of twenty-nine strategies can offer an equivalent return higher than the benchmark. Nonetheless, when the investor shows a higher degree of risk aversion, only one strategy, the book-to-market (BM) rate, can beat the historical average.

In the case of an investor with a CRRA=1, the utility gain obtained for a portfolio constructed with the historical average is negative and equal to -0.2%. It has a 0 out-of-sample Sharpe ratio, and the turnover ratio is 0%. It means that the benchmark portfolio invested all the weight in the risk-free asset initially and did not rebalance the portfolio for the whole out-of-sample period. This behaviour explains the benchmark strategy's negative certainty return because the risk-free investment, measured as the Euribor 1M yield, has yielded negative rates since February 2015. A similar pattern can be observed for some valuation ratios. In particular, portfolios constructed with dividend price rates (DP), dividend yields (DY), and earning prices (EP), are not rebalanced and do not invest in equities for the whole period either.

Regarding economic predictors, only five of the twelve variables produce economic value for a more tolerant risk-averse investor. Book-to-market (BM), short (EUR3M), long term interest rates (SWAP10Y), the yield curve (TMS), and inflation (INF) show the ability to generate portfolios that beat the historical average portfolio. The BM would be the variable that reports the best performance, beating the benchmark in more than 800 basis points and showing the best Sharpe ratio of all the strategies.

Technical predictors show similar behaviour to economic ones. Nine out of fourteen variables display positive relative CER, and OBV312 is the predictor that provides the most value with the highest CER and Sharpe ratio. Comparing turnover ratios between the best economic and technical indicators, later ones give more trading signals because technical predictors exhibit higher turnover ratios than economic ones.

Finally, two of the three strategies built with principal components, F ECO and ALL, generate higher relative CER and show positive Sharpe ratios. The inclusion of technical predictors in ALL leads to higher turnover ratios and lower CER than F ECO. F TECH CER is slightly lower than the historical average, and this more unsatisfactory performance could be partly explained by the high turnover ratios displayed by technical indicators.

1.5 Summary of results

Table 5 summarises all the results obtained in previous sections. The most important conclusions are as follows.

First, there is only one economic predictor, BM, which shows the ability to forecast in-sample, out-of-sample, and produce higher equivalent returns than the historical average for investors with different risk aversion levels. Only when investors are more tolerant to riskier strategies (CRRA=1), other predictors such as EUR3M, SWAP10Y, INF, or FECO predict better than the benchmark and produce economic value.

Table 5: Summary of in-sample, out-of-sample and asset allocation results

	In-Sample		Out-of-Sample		Asset Allocation		
Predictor	**t-stat**	**R2 (%)**	**t(MSFE)**	**R2OS (%)**	**Rel CER (CRRA=1)**	**Rel CER (CRRA=5)**	**Rel CER (CRRA=10)**
DP	0.2	0,9	0.7	0,1	-0,2	-0,2	-0,2
DY	0.3	0,7	1.0	0,3	-0,2	-0,2	-0,2
EP	0.1	0,9	0.3	0,1	-0,2	-0,2	-0,2
DE	0.7	0,5	1.2	0,6	-0,9	-0,4	-0,3
RVOL	0.	1,0	-0.6	-0,3	-0,4	-0,3	-0,2
BM	**2.6*** **	**0,0**	**3.4*** **	**4,5**	**8,3**	**1,7**	**0,8**
EUR3M	**-3.1*** **	**0,0**	**1.7***	**-3,7**	**5,0**	-7,2	-5,8
SWAP10Y	**-2.4** **	**0,0**	**1.6***	**-5,5**	**5,0**	-7,3	-7,1
TMS	**1.7***	**0,1**	-0.8	-4,5	**0,8**	0,0	-0,1
DFY	-1.5	0,2	1.1	0,6	-3,0	-1,8	-1,0
DFR	-1.4	0,2	1.2	1,1	-3,7	-1,5	-0,9
INF	**-3.2*** **	**0,0**	**1.7***	**-1,7**	**3,2**	-1,2	-3,3

	In-Sample		Out-of-Sample		Asset Allocation		
Predictor	**F-stat**	**R2 (%)**	**t(MSFE)**	**R2OS (%)**	**Rel CER (CRRA=1)**	**Rel CER (CRRA=5)**	**Rel CER (CRRA=10)**
FECO	**5.1** **	**5,8**	**1.9** **	**-2,6**	**5,0**	-3,2	-3,9
FTECH	**3.1** **	**2,5**	-0,1	-7,0	-0,3	-2,2	-1,2
FALL	**5.7*** **	**6,5**	**1.4***	**-2,8**	**0,4**	-0,7	-0,6

	In-Sample		Out-of-Sample		Asset Allocation		
Predictor	**t-stat**	**R2 (%)**	**t(MSFE)**	**R2OS (%)**	**Rel CER (CRRA=1)**	**Rel CER (CRRA=5)**	**Rel CER (CRRA=10)**
MA1MA9	**2.0** **	**0,0**	-0,3	-6,1	**1,6**	-1,1	-0,6
MA1MA12	**1.9** **	**0,1**	0,0	-4,8	**2,1**	-0,9	-0,6
MA2MA9	**1.9** **	**0,1**	0,0	-3,8	**2,4**	-0,5	-0,4
MA2MA12	1.0	0,3	-0,9	-5,7	-3,9	-2,0	-1,1
MA3MA9	**2.0** **	**0,0**	0,2	-3,3	2,6	-1,2	-0,7
MA3MA12	1.0	0,3	-0,3	-3,3	-2,1	-1,4	-0,8
MOM9	**1.9** **	**0,1**	0,2	-3,8	**2,1**	-1,4	-0,8
MOM12	0.6	0,5	-0,4	-2,4	-4,0	-1,2	-0,7
OBV19	1.1	0,3	-0,8	-5,1	-3,8	-1,9	-1,1
OBV112	**2.0** **	**0,0**	-0,2	-6,7	-0,3	-2,9	-1,6
OBV29	**2.0** **	**0,0**	0,0	-5,7	**0,7**	-2,3	-1,3
OBV212	**2.1** **	**0,0**	0,1	-5,5	**1,0**	-2,4	-1,3
OBV39	**2.5** **	**0,0**	0,6	-3,1	**3,2**	-1,7	-0,9
OBV312	**2.9*** **	**0,0**	1,3	-0,7	**6,0**	-1,0	-0,6

, ** and * indicate significance at the 10%, 5% and 1% levels, respectively. Based on two-tail bootstrapped p-values for in-sample estimations, and on Clark and West (2007)* MSFE-Adjusted for out-of-sample predictions.
R2(%) is the R^2 obtained from ols estimations, and R2OS(%) is the Campbell and Thompson (2008) R^2.
Rel CER is the relative CER calculated as the difference between the CER generated by a predictor and the CER generated by the benchmark.
CRRA would be the risk aversion coefficient, and CER includes a 50bps trading costs for each trade realized.

Second, similarly to Neely et al. (2014) for the US, we find that technical indicators and multivariate analysis using PCA gathers relevant information and show the ability to forecast in-sample. Nonetheless, we cannot confirm their results out-of-sample. Contrary to them, we do not find evidence suggesting that technical variables predict better than the benchmark out-of-sample. Neither at the individual level of each technical indicator nor grouping the most relevant information through principal components.

Third, this exercise finds that multivariate analysis using PCA gathers relevant information and creates economic value only for strategies that combine economic (F ECO) or economic and technical factors (ALL), but not only technical ones (F TECH). Technical variables seem

to rebalance more frequently the portfolio, increasing trading costs and reducing portfolio returns.

Forth, macroeconomic indicators seem to perform better than valuation ratios within the economic variables, except for BM. Further studies using more extended forecasting periods might change this result and improve valuation ratios forecasting power.

1.6 Conclusions

This chapter analyses whether traditional economic predictors and common technical indicators show any ability to forecast the equity risk premium in the EMU area. The studied period spans from January 2000 to December 2020. This study focusing on the EMU area is of interest because it further complements other papers on US data by creating a new set of results for a different market and period.

Our in-sample outcomes align with other papers on US data, and technical indicators show a better ability to forecast equity risk premiums than economic ones. Our results also match because multivariate regressions built using PCA gather relevant information from all individual predictors and improve in-sample forecasting power. We also find that economic predictors such as the nominal interest rates (EUR3M, SWAP10Y), inflation (INF), or book values (BM) display in-sample and out-of-sample forecasting ability. Besides, they provide economic value for a low risk-averse investor who invests her wealth between risk-free assets and equities. Nonetheless, our out-of-sample and asset allocation exercises do not confirm what we found in-sample and implied substantial departures from Neely et al. (2014) results for the US. First, technical predictors do not show out-of-sample forecasting ability at all. Neither at the individual level nor by grouping them with principal components. In addition, almost none of the economic or technical predictors add economic value to an investor with a medium (CRRA=5) or high (CRRA=10) degree of risk aversion. Only the BM shows the ability to forecast in-sample, out-of-sample and produces higher equivalent returns than the historical average for investors with several risk aversion levels.

These results are relevant for practitioners as they shed more light on the identity of economic and technical variables that can be useful to forecast equity markets and answers if similar trading rules can be used across geographies in a more globalised world. Moreover, we envisage four potential directions for future research that would complement and build on the

present study. First, it would be interesting to work with more extended time frequencies (e.g. quarterly or annual data) to see if results obtained at monthly frequency are consistent. Second, we worked with literature variables, but we could expand the number of variables and work with the evolving set of variables that finance practitioners use. For example, with the advent of big data, practitioners are finding innovative ways to categorise many economic and financial variables into new measures of sentiment, risk, monetary, macroeconomic, valuation or technical variables. We could improve the forecasting power of these new variables that investors and portfolio managers look at by grouping them optimally. A third way to complement this paper would be to work with time-varying parameter models. Data-generating process for stock returns can be subject to parameter instability. Models that assume estimation parameters can take different values as the economy switches between economic regimes could improve models predictability power. In this line, Markov Switching models or Dynamic Model Averaging (DMA) and Dynamic Model Selection (DMS), are exciting solutions to investigate further. To conclude, another path to explore in stock return forecasting and asset allocation is machine learning models.

References

Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? *Review of Financial Studies 20, 651-707.*

Baker, M., Wurgler, J. 2000. The equity share in new issues and aggregate stock returns. *Journal of Finance*, *55* (5), *2219–57.*

Ball, R., 1978. Anomalies in Relationship Between Securities' Yields and Yield-Surrogates. *Journal of Financial Economics, 6 (2/3)*, 103–26

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn. *Review of Financial Studies 12, 405-428.*

Boudoukh, J., Michaely, R., Richardson, MP., Roberts, MR. 2007. On the importance of measuring payout yield: implications for empirical asset pricing. *Journal of Finance 62* (2), *877–915.*

Breen, W., Glosten, LR., Jagannathan, R. 1989. Economic Significance of Predictable Variations in Stock Index Returns. *Journal of Finance 64*, *1177–89.*

Brock, W., Lakonishok, J., LeBaron, B., 1992, Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance 47*, *1731–1764.*

Campbell, JY., 1987. Stock returns and the Term Structure. *Journal of Financial Economics*, *18* (2), *373–99.*

Campbell, SD., Diebold, FX., 2009. Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence. *Journal of Business and Economic Statistics 27*, 266–278.

Campbell, JY., Shiller, RJ., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1, 195-228.*

Campbell, JY., Shiller, RJ., 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance 43, 661-676.*

Campbell, JY., Shiller, RJ., 1998. Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, *24 (2), 11–26.*

Campbell, JY., Thompson, SB., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average. *Review of Financial Studies 21(4), 1509-1531.*

Campbell, JY., Viceira, LM., 2002. Strategic Asset Allocation: Portfolio Choice for Long-term Investors, Oxford University Press, Oxford.

Campbell, JY. Vuolteenaho,T., 2004. Inflation illusion and stock prices. *American Economic Review 94, 19-23*.

Cespa, G., Vives, X., 2012. Dynamic trading and asset prices: Keynes Vs. Hyek. *Review of Financial Studies* 79, 539-580.

Clark, T., McCracken, M., 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105, 85-110.*

Clark, TE., West, KD., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138, 291-311.*

Cochrane, JH., 2008. The dog that did not bark: a defense of return predictability. *Review of Financial Studies 21, 1533-1575.*

Cochrane, JH., 2011. Presidential address: Discount rates. *Journal of Finance 66*, *1047–108.*

Cooper, I, Priestley, R., 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies 22 (7), 2801–2833.*

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106(1), 157-181.*

DeMiguel, V., Garlappi, L., Nogales, FJ., Uppal, R., 2009. A generalised approach to portfolio optimisation: Improving performance by constraining portfolio norms. *Management Science 55, 798-812.*

Dow, CH., 1920. Scientific Stock Speculation. The Magazine of Wall Street.

Driesprong, G., Jacobsen, B., Maat, B., 2008. Striking oil: another puzzle? *Journal of Financial Economics 89 (2), 307−327.*

Fama, EF., French, KR., 1988. Dividend yields and expected stock returns. *Journal of Financial Economics, 22 (1), 3–25.*

Fama, EF., French, KR., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25 (1), 23–49.*

Fama, EF., Schwert, GW. 1977. Asset Returns and Inflation. *Journal of Financial Economics 5*, 115–46.

Ferreira, MA., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics 100, 514-537.*

Goyal, A., Welch, I., 2003. Predicting the Equity Premium with Dividend Ratios. *Management Science 49, 639–54.*

Granville, JE., 1963. Granville's New Key to Stock Market Profits. Prentice Hall, New York

Guidolin, M., Timmermann, A., 2007. Asset Allocation under multivariate regime switching. *Journal of Economic Dynamics and Control 31, 3503-3544.*

Guo, H., 2006. On the out-of-sample-predictibility of stock market returns. *Journal of Business 79, 645-670.*

Han, Y., Yang, K., Zhou, G., 2013. A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative Analysis 48, 1433–1461.*

Jegadeesh, N., Titman, S. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance 48, 65–91.*

Kothari, S., Shanken, J. 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. *Journal of Financial Economics*, *44 (2), 169–203.*

Lamont, O., 1998. Earnings and expected returns. *Journal of Finance 53 (5), 1563–87.*

Lettau, M., Ludvigson, S. 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance 56 (3), 815–49.*

Lettau, M., Van Nieuwerburgh, S.. 2008. Reconciling the Return Predictability Evidence. *Review of Financial Studies 21, 1607–1652.*

Lo, A., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance 55, 1705–1770.*

Neely, J., Rapach DE., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science 60, 1772-1791.*

Nelson, CR. 1976. Inflation and the Rates of Return on Common Stock. *Journal of Finance 31, 471–83.*

Pástor, L., Stambaugh, RF. 2009. Predictive Systems: Living with Imperfect Predictors. ´ *Journal of Finance 64 (4), 1583-628.*

Pontiff, J., Schall, LD. 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics 49 (2), 141–60.*

Rangvid, J. 2006. Output and Expected Returns. *Journal of Financial Economics 81*, 595–624.

Rapach, DE., Strauss, JK., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to real economy. *Review of Financial Studies 23(2), 821-862.*

Rapach, DE., Zhou, G., 2013. Forecasting Stock Returns. Elliot, G., Timmermann, A., editions. *Handbook of forecasting, Vol 2A (Elsevier, Amsterdam), 328-383.*

Rozeff, MS., 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management 11, 68-75.*

Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. *Review of Financial Studies 19, 1–44.*

Stambaugh, RF., 1999. Predictive regressions. *Journal of Financial Econometrics 54, 375-421.*

Theil, H., 1971. Applied Economic Forecasting. North-Holland, Amsterdam.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21, 1455-1508.*

Appendix 1. Literature review

Table A1: A review of the literature. Equity risk premium predictors

Studied Variables	**Article**
Dividend-Yield ratio	Dow (1920) Ball (1978) Rozeff (1984) Campbell (1987) Campbell and Shiller (1988a,1998) Fama and French (1988, 1989) Campbell and Viceira (2002) Cochrane (2008, 2011) Lettau and Van Nieuwerburgh (2008) Pástor and Stambaugh (2009)
Earnings-Price ratio	Campbell and Shiller (1988b,1998)
Book-to-Market ratio	Kothari and Shanken (1997) Pontiff and Schall (1998)
Nominal interest rates	Fama and Schwert (1977) Campbell (1987) Breen et al. (1989) Ang and Bekaert (2007)
Interest rate spread	Campbell (1987) Fama and French (1989)
Inflation	Nelson (1976) Campbell and Vuolteenaho (2004)
Dividend-payout ratio	Lamont (1998)
Corporate issuing activity	Baker and Wurgler (2000) Boudoukh et al. (2007)
Consumption-wealth ratio	Lettau and Ludvingson (2001)
Stock market volatility	Guo (2006)
Labour income	Santos and Veronesi (2006)
Aggregate Output	Rangvid (2006)
Output gap	Cooper and Priestly (2009)
Expected Business Conditions	Campbell and Diebold (2009)
Oil prices	Driesprong et al. (2008)
Price Moving Averages	Brock et al. (1992) Lo et al. (2000) Han, et al. (2013)
Price Momentum	Jegadeesh and Titman (1993)
Multiple economic variables	Welch and Goyal (2008) Campbell and Thompson (2008) Rapach and Zhou (2013)
Multiple economic and technical variables	Neely et al. (2014)

Chapter 2

Forecasting the European Monetary Union equity risk premium with regression trees

2.1 Introduction

Forecasting stock returns is one of the most popular themes for both academics and practitioners in finance. Rapach and Zhou (2013) offer a profound revision of the available research in this area. The existing literature has studied many types of variables and proposed multiple econometric models to see if there is significant evidence of returns predictability. It is generally accepted among financial economists that stock returns contain a significant predictable component in-sample. For instance, Rozeff (1984), Campbell and Shiller (1988a) or Cochrane (2008) find evidence in favour of return predictability using the dividend yield, and Campbell and Shiller (1988b,1998) use the earnings-price ratio and reach similar results. However, there is no such consensus when forecasting ability is studied out-of-sample. Bossaerts and Hillion (1999), Goyal and Welch (2003) and Welch and Goyal (2008) show that a long list of predictors from the literature cannot perform consistently better out-of-sample than a simple forecast based on the historical average.

A big part of the literature that tries to find out-of-sample evidence of the equity risk premium predictability has focused on parametric models. For instance, Welch and Goyal (2008) or Neely et al. (2014) use univariate regression analysis to find whether traditional economic and technical predictors display out-of-sample forecasting ability. Rapach et al. (2010) illustrate that combining several individual forecasts produces robust out-of-sample forecasts. In the same line, Dangl and Halling (2012) look at Bayesian models, which assume estimation parameters can take different values as the economy switches regimes.

Nonetheless, little attention had been given to statistical learning algorithms. Few exceptions are Cao and Tay (2001, 2003), Kim (2003), Huang et al. (2005), Wang and Zhu (2010) and Miller et al. (2015), who tried support vector machine (SVM) models to forecast the financial markets. Nonetheless, with the explosion of big data, some of these learning algorithms have become very popular for constructing prediction models, and researchers have begun to use these techniques to see if they can contribute to making better predictions of the equity risk premium. On this line, Coqueret and Guida (2018) build regression trees to determine which firm characteristics are the most likely to drive stock returns. Wolf and Neugebauer (2019) use tree-based machine learning approaches to forecast the S&P 500 risk premium and find that regression trees fail to outperform the benchmark out-of-sample. Nonetheless, they discover that tree algorithms can create economic value for investors since tree trading strategies outperform a passive buy and hold investment. Gu et al. (2020) conduct

a comparative analysis of machine learning methods for finance. They demonstrate significant economic gains to investors using trees and neural networks to forecast the equity risk premium in the US.

This article works with Classification and Regression Trees (CART) methods to forecast the equity risk premium. CART were first introduced by Breiman et al. (1984), and these non-parametric models show several exciting advantages over other learning algorithms and parametric techniques. One of the benefits is their simplicity to implement and interpret results. Trees are built recursively, partitioning the original data set and grouping observations with similar response values. Final results are summarised in a tree structure, easy to interpret, determining the proper hierarchy and interaction of all independent variables. Another advantage lies in the fact that it is a non-parametric model and, hence, there are no implicit assumptions underlying relationships between the predictor variables and the dependent variable. It also implies that CART methods can be used even when there is a non-linear relationship between explained and explicative variables. Finally, another great advantage is that CART models can work with plenty of explicative variables giving consistent results. Yet, like any other statistical model, CART show some limitations. Prediction results provided by regression trees are pretty unstable. This means that small changes in sample data could prompt very different results. A solution to this instability problem is to apply ensemble methods that create multiple trees through bootstrapping techniques or any other iteration process and ensemble all results obtained. Aggregating numerous predictions will reduce variability and give more stable results.

The paper investigates the capacity of three regression trees ensemble methods (bagging, random forests and boosting) to select good economic predictors and improve the out-of-sample forecast of the equity risk premium. Our sample covers European Monetary Union (EMU) monthly data from January 2000 to December 2020. We work with an aggregated European equity index to measure the equity risk premium, the MSCI EMU. As explanatory variables, we include multiple economic and financial variables often used by practitioners in finance, such as valuation ratios, technical measures of traded volumes and prices, or economic confidence surveys. Specifically, we selected 26 market variables measured in levels and monthly changes and generate 52 independent predictors that are all loaded into the algorithms.

In line with the literature (see, for example, Welch and Goyal (2008), Campbell and Thompsom (2008) or Ferreira and Santa Clara (2011)), we compare the predictive accuracy of

regression trees to a benchmark: the historical mean average. Forecasts performance is analysed in terms of the Campbell and Thompson R^2 (R^2_{os}), which compares the MSFE of regressions constructed with selected predictors against the MSFE of the benchmark. Besides, regression trees forecasting power is also compared to simple univariate regressions. Forecasts are calculated monthly. Once a one-period equity risk premium is predicted, the period's data is added to the estimation window, making the estimation window grow period after period. Our results suggest that regression trees do not show more out-of-sample forecasting accuracy than a naïve benchmark, similar to the outcomes obtained by Wolff and Neugebauer (2019), but contrary to those found by Gu et al. (2020), both with US data.

To conclude, we measure any economic value in the out-of-sample predictions for a risk-averse investor with a quadratic utility function. We maximise investors' utility function using Brandt and Santa-Clara (2006) method of dynamic portfolio selection, assuming the investor either invests in equities or the risk-free asset. We selected this approach because it introduces economic variables in the asset allocation process. The optimisation considers that if some predictor variables gather economic conditions and contribute to predicting equity risk premiums, these predictor variables, known as state variables, should be deemed to form optimal portfolio weights. The outcomes we obtain in this section are mixed because depending on which performance measure is selected, it could be stated that regression tree algorithms can create economic value or not.

Our study contributes to the equity risk premium literature in two ways. Firstly, it adds to the growing literature using machine learning techniques, but it focuses on a European dataset. Secondly, our results raise questions about machine learning algorithms' suitability when the data set dimension is not very high.

The remainder of this study is organised as follows. Section 2.2 introduces classification and regression trees (CART). Section 2.3 focuses on the empirical analysis of the data, fitting first regression trees and analysing which predictors contribute the most to explain the equity risk premium. Later, it compares out-of-sample results between regression trees and univariate linear regressions. Section 2.4 reports a final exercise where we measure economic value for a risk-averse investor by incorporating equity risk premium predictors into its optimal asset allocation. Finally, the last section concludes.

2.2 Classification and Regression Trees (CART)

CART models were introduced by Leo Breiman, Jerome Friedman, Richard Olshen and Charles Stone in 1984 in their book "Classification and Regression Trees". CART is a non-parametric modelling technique that fixes a set of rules upon the explanatory variables to classify the explained variable into categories. It looks at the data set variables, determines which are the most important ones, and results in a decisions' tree that best partitions the data. The principal idea behind a decision tree is to recursively partition the space into smaller sub-spaces where similar response values are grouped. After the separation is completed, a constant value of the response variable is predicted within each area. The main difference between classification and regression trees is that, in the former case, the dependent variable is categorical, and the tree is used to identify the class within which this dependent variable is more likely to fall. On the contrary, regression trees are used when the dependent variable is continuous, and the tree is used to estimate its value.

The typical structure of a CART is illustrated in Figure 1. Every tree comprises at least one node, and every node corresponds to a region in the original space. Node A, the root node, gathers all the original data, whilst B2, a child node of A, results from splitting A into two different regions of the original space, B1 and B2. Child nodes together always occupy the same region of the parent node. Moreover, there is always a terminal node in every tree, also known as the leaf node. A terminal node cannot be split further, and every leaf is assigned with a unique class or value. In Figure 1, B2, C1 and C2 are the terminal nodes.

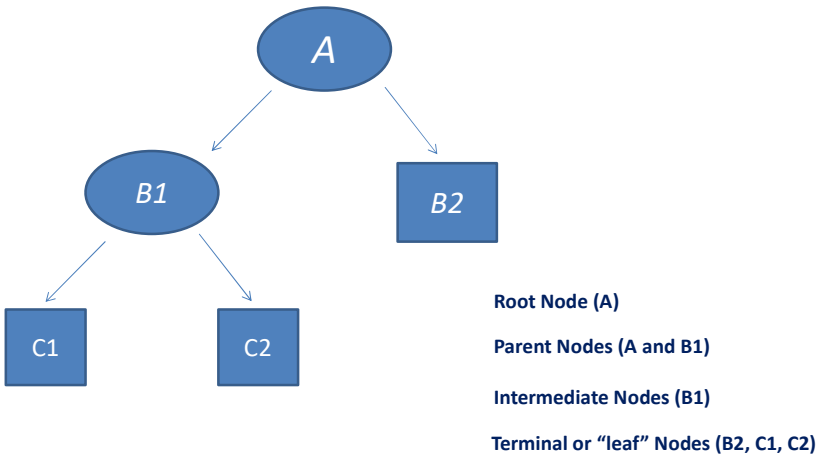


Figure 1: Example of a classification tree structure

2.2.1 Constructing the tree

The construction of a CART involves three basic steps:

1. *Start with an empty tree.*
2. *Select a feature to split data.* Tree-structured classifiers are constructed by doing repetitive splits of space X, so that a hierarchical structure is formed. The initial space X is divided into smaller subspaces, and the splitting process involves a set of binary questions of the form {is x Є Bj?}. It means asking whether the input x belongs to a certain region B. For example, initial space X could be divided into subgroups B1={x|$x_2 \leq 50$} and B2={x|$x_2 > 50$}. Then the first of these subgroups could be further divided into C1= {x|$x_2 \leq 50$, $x_1 \leq 10$} and C2= {x|$x_2 \leq 50$, $x_1 > 10$}, and the other subgroup could be split into C3= {x|$x_2 > 50$, $x_1 \leq 25$} and C4= {x|$x_2 > 50$, $x_1 > 25$}. Figure 2 illustrates this example.

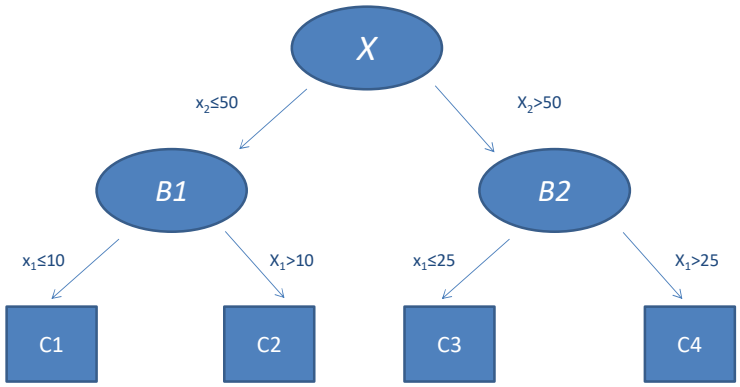


Figure 2: Partition of space X in different subregions

CART splitting keeps things simple because every split depends on the value of only a single variable. For example, if we have a numerical variable x_j, the set of questions includes all the questions of the form:

$$\{\text{Is } x_j \leq c?\} \quad \text{for all real-valued } c$$

Or if x_j was a categorical variable, then the set of questions would include questions of the form:

$$\{\text{Is } x_j \in A?\} \quad \text{where A is any subset of } \{1,2,\ldots,M\}$$

This process, known as "top-down greedy recursive partitioning", follows a top-down approach, and it is applied at every node and for every k variables. It begins at the top of the tree, includes all the data, and then successively splits the predictor space. Each split is indicated via two new branches further down on the tree. The process is considered "greedy" because, at each step, the best division is made at that particular step, without considering whether those choices remain optimal in future stages. Hence, the algorithm does not find a globally optimal tree.

Every candidate split is evaluated using a goodness measure which can be assessed for any split at any node (t). This "goodness of the split" is measured through an impurity function $\Phi(s,t)$. This impurity function measures the extent of purity for a region containing data points from possibly different classes. To select the splitting variable and cutting points, CART follows the approach of impurity reduction. The variable and cut point, which generate the highest impurity reduction, are the selected ones.

In summary, to perform recursive binary splitting, we first select the predictor x_j and cutting points, such that splitting the predictor space into the regions $\{x|x_j<s\}$ and $\{x|x_j>s\}$ leads to the greatest possible impurity reduction. We consider all predictors x_1, x_2, …, x_p, and all possible values of the cutting points for each of the predictors and then choose the predictor that maximises the impurity reduction.

3. *Stop splitting and decide which nodes are terminal.* The trickiest part of creating a good tree-structured classifier is determining how complex the tree should be. A too-large tree may produce good predictions on the training set, but it would probably "overfit" the data. The term overfitting refers to the fact that a classifier that adapts too closely to the learning sample will discover not only the systematic components of the structure that are present in the population, but also the random variation from this structure that is present in the learning data.

   Each node splitting would continue recursively until some stop condition is reached and the grown tree has a suitable complexity. There are several stop criteria, but one of the best approaches is proposed in Breiman et al. (1984) and consists of letting the tree grow to saturation and then prune it back[1].

[1] See Appendix 1 for a more detailed explanation of the pruning process.

2.2.2 Implementing CART to create a regression tree

We talk about regression trees when the dependent variable is numerical instead of categorical. The whole process to create trees is very similar, and the main difference between building classification or a regression tree lies in how the predictor space is split and how forecasted values, instead of categories, are assigned to each node.

Starting with the latter difference, in regression trees for every observation that falls into a node or sub-region of the data, R_j, the same predicted value is always given. It could be the mean of observations that fall in that sub-region R_j, the median or any other descriptive statistic. In this article, we will use the mean as the single predicted value for every tree node.

Regression trees split the space and construct new nodes minimising the residual sum of squares (RSS) regarding recursive splitting. Hence, we select the predictor X_j and cut points such that splitting the predictor space into subsets $\{X|X_j < s\}$ and $\{X|X_j \geq s\}$ leads to the greatest possible reduction in the RSS. The goal is:

$$\min \sum_{j=1}^{J} \sum_{i \in R_{(j,s)}} \left(y_i - \hat{y}_{R_j}\right)^2 \quad (1)$$

Where $\hat{y}_{R_j}$ is the mean response for the training observations within the *j*th subset of data.

2.2.3 Problems with trees

A problem with classification and regression tree models is their instability to small changes in the learning data. It happens because, in recursive partitioning, the exact position of each cut point in the partition and the decision of which variable to split in determines how the observations are divided in the subsequent nodes (bear in mind that CART use a top-down greedy recursive partitioning). For this reason, trees are highly unstable, and their structure could change dramatically if early splitting variables and cutting points are shifted due to small changes in the training set. It makes decision trees suffer high variability.

One possible solution to this instability problem would be to generate multiple trees through bootstrapping, or any other statistical technique that creates numerous training sets, gets separate predictions for each tree, and aggregate those predictions. These ensemble

methods aim to improve the predictive performance and reduce model variability using a combination of regression trees instead of a single one.

$$\hat{f}_{ensemble}(x) = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t(x) \quad (2)$$

The most common types of ensembles used with trees are bagging, random forest and boosting techniques. Breiman (1996a, 1996b and 2001) initially proposed bagging and random forest models and both techniques bootstrap the original training set, get multiple trees and average results. The difference between boosting and random forest lies exclusively in the number of predictors considered for each split in the trees. Bagging techniques consider all predictors at each break for every tree that is created through bootstrapping. Nonetheless, the random forest does not use all the explicative variables at each split, but a subset of these. Random forests improve bagged trees because simply selecting a random subset of predictors de-correlates the resulting trees. Thus, for instance, suppose that there is a powerful predictor that dominates all the others. Then, in creating bagged trees, this predictor would be on almost all trees' top splits. This would not happen if a random selection of the predictors were chosen.

On the other hand, boosting is an iterative procedure that does not involve bootstrap sampling. Instead, trees are grown sequentially, and every new tree is fit on a modified version of the original data set. There are many boosting algorithms, but the first ones were proposed by Freund and Schapire (1997).

A naïve formulation of the boosting techniques for regression trees can be described as follows:

1. Fit a regression tree with d splits to the training data.
$$F_1(x) = y$$
2. Fit a new regression tree with d splits to residuals obtained from the first regression tree.
$$h_1(x) = y - F_1(x)$$
3. Create a new model, including a weighted version of the new tree.
$$F_2(x) = F_1(x) + \lambda h_1(x)$$
4. Repeat the process iteratively B times until a final output of the boosted model is reached.

$$F_B(x) = F_1(x) + \sum_{b=1}^{B} \lambda\, h_b(x)$$

Hence, we fit a new regression tree to the initial model residuals given the current regression tree model. We then add this new regression tree into the initial fitted function to update the residuals' information.

The shrinkage parameter (λ) is a small positive number that generally takes a value between the interval [0,1], and controls the rate at which boosting learns.

2.3 Forecasting the equity risk premium

2.3.1 The data and the model

Classification trees are alternative non-parametric approaches that do not need model assumptions, allow to work with many variables, regardless of whether they show a strong correlation or not, and consider both linear and non-linear relations between predicted variables and predictors.

Regression trees assume a model of the form:

$$ERP_{t+1} = \sum_{m=1}^{M} c_m x \prod_{i=1}^{n} I(x_{i,t} \in R_m) \quad (3)$$

Where,

ERP_{t+1} is the equity risk premium in moment t+1.

c_m are constants.

$I(.)$ is an indicator function returning 1 if its argument is true and 0 otherwise.

$x_{i,t}$ are the predictor variables at time t.

$R_1, \ldots, R_m$ represent a partition of the feature space.

To avoid the variability problems suffered by individual trees, we focus on ensemble techniques that generate multiple trees and average all predictions produced at the individual level. These methods reduce the variance of regression forecasts, increase prediction accuracy,

and solve instability. Specifically, we estimate the three classifier methods described before: bagging, random forest and boosting.

Our empirical analysis is conducted for the European Monetary Union with a monthly dataset covering a period from January 2000 to December 2020. The equity risk premium, which is the expected return in excess of the risk-free rate, is calculated as the difference between the MSCI EMU index's monthly returns and the Euribor 1 month rate. Data is obtained from Bloomberg, Refinitiv, ICE Fixed Income Indices and Haver Data Analytics.

$$ERP_t = \left[\frac{MSCI\ EMU_t}{MSCI\ EMU_{t-1}} - 1\right] - Euribor\ 1M_{t-1}$$

Table 1 shows all the explanatory variables used to forecast the equity risk premium[2]. We grouped these predictors into five groups: Valuation/Fundamental variables, Rates & Inflation, Risk indicators, Confidence Surveys and Technical indicators. These groups contain a significant part of the variables that practitioners analyse to form their market expectations and forecast the equity risk premium.

Table 1: Predictive variables

Economic and Technical Variables				
Valuation/Fundamentals	**Rates & Inflation**	**Risk Indicators**	**Confidence Surveys**	**Technical indicators**
Dividend Yield	Euribor 3M	VIX	Consumer Confidence	Moving Averages
Price to Book Value	Swap Rate 2Y	ASW IG Corporates	EU Business Climate Indicator	OBV
Price to Earnings (trailing)	Swap Rate 10Y	ASW HY Corporates	Industrial Confidence	Earnings momentum
Price to Earnings (forward)	Yield Curve 3m-10Y	Spread IG -HY Corporates	Retail Confidence	
Price to Earnings (trailing, 5Y averages)	Yield Curve 2Y-10Y	TED Spread	OECD EU Leading Indicator	
Price to Earnings (forward, 5Y averages)	NYMEX level			
	Consumer Prices			

Variables are measured in levels and monthly changes, which gives a total of 52 predictors that are loaded into the algorithm. We include all variables because results obtained from trees are consistent regardless there is redundant information, and we are unsure whether predictors work better in levels or changes

We resample trees 500 times[3] to run bagging and random forest algorithms, considering all 54 predictors at each split, in the case of bagging, and seven randomly selected ones (approximately the squared root of the 52 variables) for random forests. Recall that bagging is

[2] See Appendix 2 for a detailed definition of the explanatory variables.

[3] We resampled the trees 500 times because it is the default number used in the R library (randomforest) to calculate regression trees. We avoided fitting the number of sample trees through cross-validation or any other fitting technique to keep things as simple as possible and avoid data mining problems.

simply a particular case of a random forest where all explicative variables are considered at each split. In boosting, we implement Freund and Shapire (1997)'s AdaBoost algorithm's extensions and Friedman (2001)'s gradient boosting machine. We fit a generalised boosted regression, with 5000 trees[4] to fit with a maximum depth of 4 for each tree.

Figure 3 lists the ten most important predictors for the three tree classifiers considering the whole sample and using the residual sum of squares as the impurity measure. Importance is calculated by looking at how each feature reduces mean squared errors at each node and then averages across trees to determine each variable's extent.

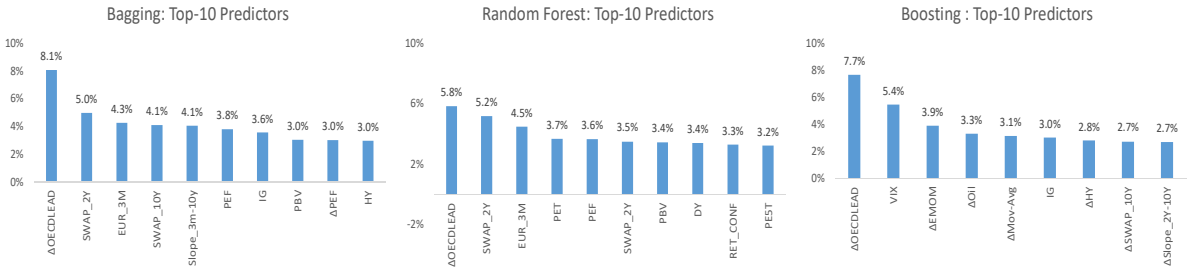


Figure 3: Bagging, random forest and boosting top-10 important variables

The three algorithms' variable selection indicates that all three models favour the monthly variation of the OECD Leading Indicator Index (ΔOECDLEAD) as the best predictor to forecast the equity risk premium. Apart from this variable, the rest of the predictors selected by the bagging and random forest models are very different from those chosen by boosting. The first two algorithms perform a similar selection of variables and share six of the ten best predictors (ΔOECDLEAD, EUR_3M, SWAP_2Y, SWAP_10Y, PEF and PBV), showing structural similarities and high correlation in their predictions. Moreover, these two processes tend to select variables in levels instead of monthly changes. On the contrary, the boosting algorithm picks a different group of variables as best predictors, and most of them in first differences.

[4] Similar to bagging and random forests algorithms, we use the default parameters given by the R library (gbm), to keep things as simple and replicable as possible.

2.3.2 Out-of-sample forecasts

To analyse the out-of-sample forecasting ability of all three models, we divide the sample into two subsets. The first one, known as the training set, spans from January 2000 to December 2012. The second, the out-of-sample subset, ranges from January 2013 to December 2020 (96 observations). We compute the predictive regression trees for the in-sample estimation period and then calculate recursively out-of-sample forecasts one period ahead. The forecasts employ an expanding window, meaning that the estimation sample always started in January 2000, and additional observations are included as they become available.

In line with the literature (see Welch and Goyal (2008), Campbell and Thompsom (2008) or Ferreira and Santa Clara (2011)), we compare the predictive accuracy of regression trees to a benchmark. The selected benchmark is the historical mean average, which assumes that its historical average will give the future value of excess returns.

$$\hat{r}_{t+1}^{HA} = \frac{1}{t}\sum_{i=1}^{t} r_i \tag{4}$$

We compare forecasts in terms of the Campbell and Thomson (2008) out-of-sample R^2 (R_{OS}^2).

$$R_{OS}^2 = 1 - \frac{MSFE_r}{MSFE_b} \tag{5}$$

The $MSFE_r$ is the mean squared forecast error of the predictive regression, whereas the $MSFE_b$ is the mean squared forecast error of the benchmark. This R_{OS}^2 measures the proportional reduction in the MSFE for the predictive regression forecast relative to historical averages. Thus, if $R_{OS}^2 > 0$, then predictive regressions forecasts relative to historical averages are more accurate. But if $R_{OS}^2 < 0$, then predictive regression forecast cannot beat the benchmark.

We also determine whether the improvement in the forecast is significant. Diebold and Mariano (1995) proposed a statistic (DM statistic) for testing the null of equal MSFE between two models. We implement the modified test offered by Harvey et al. (1997), which tests the null hypothesis that the benchmark and the selected regression have the same forecasting accuracy, H_0: $MSFE_b = MSFE_r$, and the alternative is that the regression chosen is more accurate than the benchmark, H_a: $MSFE_b > MSFE_r$.

Table 2: Out-of-sample forecasting results

					Theil MSE % Decomposition		
Predictor	MSFE	ROS	t-stat	p-value	Bias	Variance	Covariance
Benchmark	0,0021	0,0			2,4%	92,9%	4,7%
Banging	0,0023	-0,1	1,67	0.95	0,0%	45,1%	54,8%
Random Forest	0,0023	-0,1	2,00	0.97	0,1%	57,5%	42,4%
Boosting	0,0030	-0,4	3,84	0.99	0,1%	12,9%	87,0%
Top-5 Predictors							
ΔInd_conf	0,0007	0,7	-4,88	0.00***	3,9%	82,5%	13,6%
ΔSwap_2Y	0,0008	0,6	-4,01	0.00***	3,7%	88,8%	7,5%
ΔCons_conf	0,0009	0,6	-4,84	0.00***	3,7%	84,7%	11,6%
ΔRetcon	0,0010	0,5	-4,83	0.00***	3,6%	93,2%	3,1%
ΔMov-Avgs	0,0018	0,1	-5,04	0.00***	2,5%	96,1%	1,3%
Bottom-5 Predictors							
Swap_2Y	0,0063	-2,0	5,01	0.99	1,4%	2,0%	96,7%
Euribor3M	0,0066	-2,1	4,92	0.99	0,8%	1,4%	97,8%
Mov_avgs	0,0153	-6,2	14,15	0.99	88,5%	6,8%	4,7%
ΔOECD_Lead	0,0608	-27,5	4,29	0.99	0,3%	96,9%	2,8%
OECD_Lead	0,0774	-35,3	27,54	0.99	97,0%	1,3%	1,8%

, and *** indicate significance at the 10%, 5% and 1% levels, respectively.*
MSFE is the mean squared forecast error for the out-of-sample period.
ROS is the Cambell and Thompson R^2 statistic.
t-stat and p-value are the statistic and the p-value of the Diebold-Mariano test for predictive accuracy.
Univariate linear regressions are calculated as $r_{t+1} = \alpha_t + \beta_t * x_{i,t}$, where α_t and β_t are the univariate estimations up to time t, $x_{i,t}$ is the predictor i at period t, and r_{t+1} is the equity risk premium at t+1.

Table 2 exhibits forecasting results for the benchmark, the three regression trees, and the best and worst univariate[5] linear regressions calculated for each predictor variable. Outcomes suggest that none of the regression trees can beat the benchmark. All three ensemble methods show higher MSFE than the benchmark and, therefore, negative Campbell and Thomson out-of-sample R^2. Moreover, DM statistics confirm that these regression trees do not show better forecasting ability than the benchmark since the null hypothesis stating that both models have similar forecasting ability cannot be rejected. Between the three methods, random forests display lower forecasting errors, while boosting shows the highest ones.

At an individual level, best performer predictors include confidence indicators such as the Industrial Confidence (ΔInd_Conf), Consumer Confidence (ΔInd_Conf) or Retail Confidence (ΔRet_Conf), an interest rate indicator, the two years swap rate (ΔSwap_2Y), and technical indicators such as the moving averages (ΔMov_avgs). All these five predictors are in monthly growth rates and not in levels. The DM statistic rejects the null hypothesis and confirms that

[5] See appendix 3 for all the univariate variables out-of-sample results.

these univariate regressions built with these individual predictors show better forecasting ability than the benchmark.

On the contrary, four out of five of the worst performer predictors are in levels. These include the OECD Leading indicators (OECD_Lead, ΔOECD_Lead), and surprisingly, two variables that in first differences are between the best top-5 predictors, but in levels are among the worst performers: the two years swap rates (Swap_2Y) and moving averages (Mov_avgs).

It is striking that between the variables with the worst predictive capacity, it can be found the best predictor selected by the three algorithms (ΔOECD_Lead) and the three best predictors selected by the bagging and random forests models (ΔOECD_Lead, SWAP_2Y and Eur_3M). Hence, it might be stated that, in this sample, ensemble regression trees do not show better forecasting ability than a naïve benchmark, and likely, it can be explained because the selection of predictors was not the most accurate one.

Another angle to understand predictions' accuracy is to look at the Theil (1971) MSE decomposition[6]. Theory suggests that forecasts are acceptable if parts explained by the bias and the variances proportions are small, and most of the MSE is defined by the unsystematic component proportion. In this line, ensemble methods show better performance than the benchmark or even the best performer predictors. All three regression trees exhibit lower bias and variances and higher unexplained proportion (covariance). These outcomes contradict those obtained with the R^2_{OS} and raises questions about whether the MSFE, which focuses on point forecast but ignores other distribution properties, is the most appropriate method to compare forecasting performance in highly erratic time series.

2.4 Asset allocation

As a final exercise, we measure whether top predictors selected by regression trees can produce economic value for a risk-averse investor with a quadratic utility function that invests her wealth combining equity with the risk-free asset. Following Brandt and Santa Clara (2006), we estimate optimal portfolio weights considering predictor variables and creating dynamic portfolios that vary their weights as economic conditions change. Assuming that optimal

[6] The Theil (1971) MSFE decomposition is given by: a) the bias component $\left(\hat{\bar{X}} - \bar{Y}\right)$, that represents how far is the mean of the forecast from the mean of the actual series; b) the variance component $(S_{\hat{X}} - S_X)$, which measures how far the variance of the forecast is from the actual series; and c) the covariance component $2(1 - r_{\hat{X}X})S_{\hat{X}}S_X$, where r is the correlation, and this component measures the remaining unsystematic forecasting error. If the forecast is good, the bias and the variance proportions should be small so that most of the bias should be concentrated on the covariance proportion.

portfolio weights are linear functions, the state variables z_t contribute linearly to calculate portfolio weights.

$$w_t = \theta z_t \tag{6}$$

Where θ is a *NxK* matrix of coefficients, and the optimal portfolio weights are linear functions of *K* state variables z_t, and the investor maximises the following expected quadratic utility function:

$$Max\ E_t\left[r_{t+1}^p - \frac{\gamma}{2}\left(r_{t+1}^p\right)^2\right] \tag{7}$$

Where r_{t+1}^p is the return on the investor's portfolio over the next period, and γ is a positive constant that represents an investor's coefficient of relative risk aversion.

Denoting the vector of portfolio weights on the risky assets at time t by w_t, and assuming that $w_t = \theta z_t$, the above optimisation problem then becomes

$$Max\ E_t\left[(\theta z_t)^T r_{t+1} - \frac{\gamma}{2}\left((\theta z_t)^T r_{t+1} r_{t+1} x_t (\theta z_t)\right)^2\right] \tag{8}$$

Where r_{t+1} is the vector of excess returns on the *N* risky assets, and using some linear algebra

$$(\theta z_t)^T r_{t+1} = z_t^T \theta^T r_{t+1} = vec(\theta)^T (z_t \otimes r_{t+1}) \tag{9}$$

Where $Vec(\theta)$ gathers the columns of θ and $\otimes$ is the Kronecker product of two matrices, and:

$$\widetilde{w} = Vec(\theta) \tag{10}$$

$$\tilde{r}_{t+1} = z_t \otimes r_{t+1} \tag{11}$$

The investor's optimisation problem can then be written as:

$$Max\ E_t[(\widetilde{w})^T r_{t+1}] - \frac{\gamma}{2} var[(\widetilde{w})^T r_{t+1}] \tag{12}$$

And the optimal portfolio weights calculated as:

$$\widetilde{w}^* = \frac{1}{\gamma}\left(\frac{E[\tilde{r}_{t+1}]}{var[\tilde{r}_{t+1}]}\right) \tag{13}$$

Once the investor knows her optimal weights between the risky asset ($\widetilde{w}^*$) and the risk-free asset ($1 - \widetilde{w}^*$), then she can calculate portfolio returns at the end of each period as:

$$r_{p,t+1} = \widetilde{w}_t^* r_{t+1} + r_f \tag{14}$$

Where r_f denotes the risk-free return from time t to t+1, r_{t+1} is the equity risk premium for the period t+1, $\widetilde{w}_t^*$ is the given weight by each strategy to the equity index at period t, and $r_{p,t+1}$ is the portfolio return at period t+1.

Whether trading strategies have economic value or not will be determined by the investor's relative utility function or relative certainty equivalent return (CER)[7]. If the utility value produced by regression trees trading strategies is higher than the utility generated by the benchmark, algorithms strategies create economic value for a risk-averse investor. Otherwise, the opposite is true.

In line with De Miguel et al. (2009) and Neely et al. (2014), we also use two other criteria along with CER to compare the out-of-sample performance of the different trading strategies:

2) The out-of-sample Sharpe ratio, defined as the sample mean of the out-of-sample excess returns, $\hat{\mu}_p$, divided by their sample standard deviation, $\hat{\sigma}_p$.

$$\widehat{SR} = \frac{\hat{\mu}_p}{\hat{\sigma}_p} \tag{15}$$

3) The out-of-sample portfolio turnover. It gives the amount of trading required to implement each portfolio strategy, and it is defined as the average sum of the absolute value of trades realised.

$$Turnover = \frac{1}{T-1}\sum_{t=\tau}^{T-1}\left(\left|w_{t+1}^i - w_t^i\right|\right) \tag{16}$$

Where w_t^i is the portfolio weight in the risky asset before rebalancing for strategy i, w_{t+1}^i is the rebalanced portfolio weight in equities at t+1 for strategy i, and T is the total number of portfolio returns obtained in the estimation period.

[7] The utility function can also be interpreted as an approximation of the CER, which is the minimum guaranteed return that a mean-variance investor with a risk-aversion coefficient δ would consider equivalent to investing in the strategy. A higher CER means that a specific risk-averse investor needs higher compensation not to take the strategy.

Since we calculate the turnover for each trading strategy, it is possible to include trading costs into the portfolio returns, penalising in this way those strategies that realise more trading operations. Formally, we compute portfolio returns, including transaction costs as:

$$r_{p,t+1} = [w_t r_{t+1} + r_f] - [TO_t x\ trading\ costs] \quad (17)$$

Where TO_t is the turnover of each strategy at time t, and trading costs represent each individual's transaction monetary cost.

Moreover, we introduce three assumptions to run the out-of-sample asset allocation: 1) we assume that the investor uses a five-year moving window of past monthly returns for the estimation period. 2) We constrain w_t to lie between 0 and 1, which implies that short sales and leverage are not allowed. 3) We assume that the investor has a degree of risk aversion (δ or CRRA) equal to 1, 5 or 10.

We choose an estimation window of five years length (60 months) and an out-of-sample estimation period that spans from January 2013 to December 2020. Each month t, starting from t+1 we use the data in the previous 60 months to estimate regression trees with the most important predictors as state variables and calculate optimal portfolio weights in a portfolio of only one risky asset and only one period ahead dynamically managed. Then, with the optimal weights, we compute portfolio excess returns in t+1. This process is repeated by adding new data for the next period and dropping the earliest one until the end of the out-of-sample sub-sample is reached.

Table 3 provides estimates for a single period dynamically managed portfolio for the period that expands from January 2013 to December 2020. The first column in the table shows strategies used to build dynamic portfolios. These include: 1. The historical average portfolio, used as a benchmark in the previous section. 2. A naïve buy and hold strategy which invests 75% in the risk-free asset and 25% in the equity index. 3. The Top-1 Variable, which uses the best predictor selected by each regression tree algorithm as the state variable. 4. Top-2 Variables, which uses the two best ones. 5. Top-3 Variables, which uses the three best ones. Besides, the following columns in Table 3 include the CER gain, including 50bps transactions cost for each strategy, out-of-sample Sharpe ratios, and the out-of-sample portfolio turnover.

Table 3: Portfolio performance measures (January 2013 to December 2020)

Portfolio	CRRA=1		CRRA=5		CRRA=10		Sharpe Ratio
	CER Gain (TC=50bps)	Turnover (%)	CER Gain (TC=50bps)	Turnover (%)	CER Gain (TC=50bps)	Turnover (%)	
Historical Average	-0,02	0,04	0,00	0,01	0,00	0,00	0,00
Naïve Long Portfolio	1,57	0,00	1,26	0,00	0,87	0,00	0,12
Top-1 Variable							
Bagging	0,99	0,12	0,19	0,02	0,09	0,01	0,17
Random Forest	0,99	0,12	0,19	0,02	0,09	0,01	0,17
Boosting	1,20	0,18	0,24	0,04	0,12	0,02	0,15
Top-2 Variables							
Bagging	0,81	0,13	0,16	0,03	0,08	0,01	0,14
Random Forest	0,81	0,13	0,16	0,03	0,08	0,01	0,14
Boosting	1,19	0,22	0,24	0,05	0,12	0,02	0,14
Top-3 Variables							
Bagging	0,72	0,12	0,14	0,02	0,00	0,01	0,12
Random Forest	0,72	0,12	0,14	0,02	0,00	0,01	0,12
Boosting	0,82	0,25	0,22	0,06	0,11	0,03	0,09

CRRA is the Constant Relative Risk Aversion coefficient. It represents the investor's degree of risk aversion. Higher values mean higher risk aversion.
CER Gain (TC=50bps) is the Certainty Equivalent Return including transaction costs (TC). Transaction costs are calculated as TC x Turnover, where TC is the cost of a trade. Sharpe ratio is the monthly out-of-sample Sharpe ratio.
Turnover measures the % cost of rebalancing trades per month for each strategy.
Top-1 Variable means that optimal portfolio weights were calculated with algorithm's best performer predictor, Top-2, with the two best ones, and Top-3 with the three best ones. Naïve Long portfolio is built investing all the time 75% in the risk-free asset, and 25% in equity markets.

Table 3 shows mixed results if we observe one or the other performance measure. If we pay attention to CER values, the naïve strategy is the one that generates the highest economic value for risk-averse investors regardless of the level of risk aversion. In this sense, trees algorithms and the historical average portfolio underperform a buy and hold benchmark that invests 75% of the portfolio in the risk-free asset and 25% in equities. Thus, the certainty equivalent return of the naïve long-only portfolio for an investor with the lowest level of risk aversion (CRRA=1) is 1.57, well above the CER obtained with other trading strategies. And the same is true for higher levels of risk aversion, only that the utility provided by all strategies decreases at higher levels of risk aversion. Comparing tree regressions, boosting gives a higher average utility than the other two algorithms.

On the other hand, if we look at Sharpe ratios, regression trees strategies offer higher values than the naïve long-only portfolio and the historical average. Between the algorithms, bagging and random forests display better performance than boosting. Just the contrary to results suggested by the certainty equivalent return. Hence, depending on the performance measure we use, it could be stated or not that trees offer economic value for an investor. If the performance measure selected is CER, then the naïve portfolio outperforms those constructed

with trees and the historical average at any level of risk-aversion. On the contrary, if the performance measure selected is the Sharpe ratio, then regression trees can offer investors economic value.

One striking outcome in Table 3 is that calculations obtained with bagging and random forest are precisely the same in all cases, which means that both algorithms always select the same top predictors and show a high correlation in their predictions. Besides, using fewer predictors as state variables to decide optimal portfolio weights works better than using more. Perhaps, this can be explained because state variables included to calculate portfolio weights are equally weighted in the optimisation process, and likely variables with greater predictive capacity should take greater weight in equation (6).

Finally, Table 3 also suggest that boosting strategies require the highest amount of trading to be implemented. Boosting turnover ratios are the highest among all the portfolio strategies, followed by bagging and random forests. Moreover, it can be observed that the amount of trading and the turnover increase as the investor becomes less risk-averse.

2.5 Conclusions

In this chapter, we perform three regression tree ensemble methods (bagging, random forests and boosting) to establish the capacity of multiple economic and technical indicators to forecast the equity risk premium in the EMU area. Our sample includes a monthly dataset covering a period that spans from January 2000 to December 2020 and 26 predictor variables widely used by academics and practitioners. First, we analysed which predictors are the most important to drive the equity risk premium over the whole sample. Our results showed that the three algorithms select the monthly variation of the OECD Leading Indicator Index (ΔOECDLEAD) as the most relevant predictor to forecast the equity risk premium. Apart from this variable, bagging and random forests select similar predictors while boosting choose different ones.

Next, we checked if regression tree algorithms show significant out-of-sample forecasting ability. We compared forecasts accuracy in terms of the Campbell and Thomson (2008) out-of-sample R^2 (R^2_{OS}), and our findings indicate that regression trees do not show more out-of-sample forecasting accuracy than a naïve benchmark represented by the historical mean average of excess returns.

These results align with Wolff and Neugebauer (2019) but contradict those obtained by Gu et al. (2020), both for the US market. The former article, like this one, works with samples of reduced dimensions and with equity indices. Wolff and Neugebauer (2019) use 28 predictors to forecast the S&P 500 risk premium on a dataset ranging from June 1993 to December 2017. A sample size similar to the one used in this study. In contrast, Gu et al. (2020) find that regression trees offer high out-of-sample forecasting power for an extensive data sample, including 30.000 individual stocks, over 60 years, and more than 900 predictors. These differences between the articles' outcomes might suggest that regression trees and other machine learning techniques can be a helpful forecasting tool in richer datasets environments, while smaller dimension datasets might be analysed better with traditional parametric analysis.

Finally, we studied if regression tree algorithms can generate economic value for a risk-averse investor with different risk aversion levels, using Brandt and Santa-Clara (2006) conditional portfolio choice. The optimisation results are ambiguous because decision trees may or may not generate economic value for an investor, depending on the performance measure chosen. If the performance measure selected is CER, then the naïve portfolio outperforms those constructed with trees and the historical average at any level of risk-aversion. On the contrary, if the performance measure selected is the Sharpe ratio, then regression trees can offer investors economic value.

Apart from the above results, this article found other thrilling outcomes. First, the best predictor selected by all three algorithms in-sample, the monthly variation of the OECD Leading Indicator, is the worst out-of-sample predictor at the individual level. Moreover, the best predictors selected by the algorithms in-sample are not the best individual out-of-sample individual predictors. Second, bagging and random forests determine very similar predictors as the best ones to forecast the equity risk premium, despite reducing the number of predictors significantly to select at each split in the random forests. It raises questions of whether it is necessary to run both algorithms in small samples in which the explanatory variables show a high degree of collinearity, as it is the case with financial and economic variables. Third, R^2_{OS} suggest that regression trees show lower forecast ability than the benchmark. However, the Theil decomposition of the MSFE showed that tree algorithms had more accurate decomposition than the benchmark and most of the other predictors. This raises questions about whether the MSFE is the most appropriate method to compare forecasting performance in very volatile time series.

This paper's research is relevant for academics and practitioners because it contributes to the extensive literature on stock return forecasting but implementing machine learning techniques, such as regression trees ensemble methods, and covering the Euro equity market. Moreover, we envisage three potential directions for future research that would complement and build on the present study. First, to use larger cross-section datasets with large cross-sections and time dimensions. Hence, colinearity among explanatory variables decreases, and machine learning techniques become more powerful statistical tools. Second, future research should also compare other points forecasts, such as de variance or any relevant moment of the predicted variable distribution function, in addition to the MSFE. This approximation to the density forecasts could provide further information of each variable or group of variables ability to forecast. Third, to consider Brandt et al. (2009) for constructing optimal portfolios with cross-sectional characteristics. This asset allocation technique, called Parametric Portfolio Policies (PPP), performs portfolio optimisation that works well in large datasets because it produces robust portfolio weights, resulting in good out-of-sample performance. Moreover, optimisation exercises should include other investors' utility functions, for instance, the isoelastic utility function.

References

Bossaerts, P., Hillion, P., 1999. Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *Review of Financial Studies, 12(2), 405–428.*

Brandt, MW., Santa-Clara, P., 2006. Dynamic portfolio selection by augmenting the asset Space. *Journal of Finance 61, 2187-2217.*

Brandt, MW., Santa-Clara, P., Vakanov, R., 2009. Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *Review of Financial Studies*, *22(9), 3411-3447.*

Breiman, L., 1996a. Stacked Regressions. *Machine Learning, 24(1), 49-64.*

Breiman, L., 1996b. Bagging Predictors. *Machine Learning, 24(2), 123-140.*

Breiman, L. 2001. Random Forests. *Machine Learning, 45, 5–32.*

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees.* Wadsworth, Belmont, California.

Campbell, JY., Shiller, RJ. 1988a. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, *1(3), 195-228*.

Campbell, JY., Shiller, RJ., 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance* 43, 661-676.

Campbell, JY., Shiller, RJ., 1998. Valuation Ratios and the Long-Run Stock Market Outlook. *Journal of Portfolio Management, 24(2), 11-26.*

Campbell, JY., Thompson, SB., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average. *Review of Financial Studies, 21(4), 1509-1531*.

Cao L.J., Tay, F.E.H. 2001. Financial forecasting using support vector machines. *Neural Computing & Applications*, *10, 184-192*.

Cao L.J., Tay, F.E.H. 2003. Support vector machine with adaptive parameters in financial time series forecasting, *IEEE Trans. Neural Netw.*, *14(6), 1506-1518*.

Cochrane J. H., 2008. The Dog That Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies*, *21, 1533-75*.

Coqueret, G., Guida, T., 2018. Stock Returns and the Cross-Section of Characteristics: A Tree-Based Approach. SSRN: https://ssrn.com/abstract=3169773

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, *106(1), 157-181*.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?. *Review of Financial Studies, Volume 22, Issue 5.*

DeMiguel, V., Garlappi, L., Nogales, FJ., Uppal, R., 2009. A generalised approach to portfolio optimisation: Improving performance by constraining portfolio norms. *Management Science 55, 798-812.*

Diebold, F.X., Mariano, R.S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13, 253–263*.

Ferreira, MA., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics, 100, 514-537*.

Freund, Y., Schapire, R.E. 1997. A Decision-Theoretic Generalisation of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences, 55(1), 119-139*.

Friedman, J., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29(5), 1189-1232*.

Goyal, A., Welch, I., 2003. Predicting the Equity Premium With Dividend Ratios. *Management Science, 49(5), 639-654*.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies, 33(5), 2223–2273*.

Harvey, D., Leybourne, S., Newbold, P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13(2), 281-291*.

Huang, W. Nakamori, Y., Wang, S.Y., 2005, Forecasting stock market movement direction with support vector machine. *Computers & Operations Research, 32(10), 2513-2522*.

Kim, K.J, 2003. Financial time series forecasting using support vector machines. *Neurocomputing, 55 (1–2), 307-319*.

Miller, K.L., Li, H., Zhou, T.G., Giamouridis, D., 2015. A Risk-Oriented Model for Factor Timing Decisions. *Journal of Portfolio Management, 41 (3), 46-58*.

Neely, J., Rapach D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science, 60, 1772-1791*.

Rapach, D.E., Strauss, J., Zhou, G., 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies, 23(2), 821-862.*

Rapach, D.E., Zhou, G., 2013. Forecasting Stock Returns. Elliot, G., Timmermann, A., editions. *Handbook of forecasting, Vol 2A (Elsevier, Amsterdam), 328-383*.

Rozeff, M.S., 1984. Dividend Yields are Equity Risk Premiums. *Journal of Portfolio Management, Fall, 68-75*.

Theil, H., 1971. *Applied Economic Forecasting*. North-Holland, Amsterdam.

Wang, L., Zhu, J., 2010. Financial market forecasting using a two-step kernel learning method for the support vector regression. *Annals of Operations Research*, 174-(1), 103-120.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies, 21, 1455-1508*.

West, K.D. 1996, Asymptotic inference about predictive ability**.** *Econometrica*, *64, 1067-1084*.

Wolff, D., Neugebauer, U., 2019. Tree-based machine learning approaches for equity market predictions. *Journal of Asset Management, 20, 273-288*.

Appendix 1. Right sized trees via pruning

Figure A1 illustrates an example of a pruned tree. Tree T-T_{t2} was constructed pruning branch T_{t2} from tree T. This means deleting from T all descendants of T_{t2} except its root node.

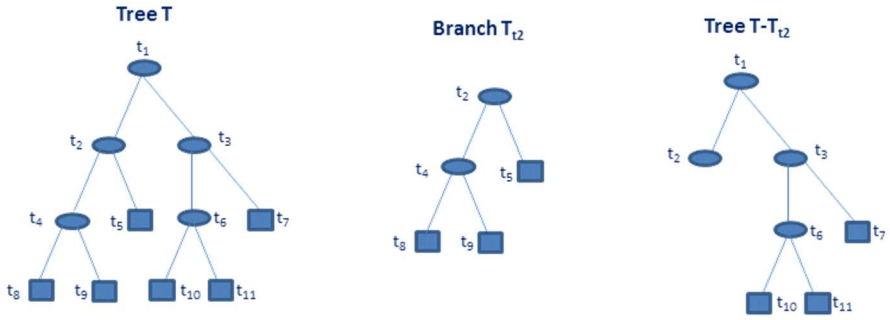


Figure A1: Example of a pruned subtree

The problem with pruning is that even for a moderated sized tree, there is an enormously large number of subtrees and an even larger number of ways to prune the initial tree. Therefore we cannot go exhaustively through all the subtrees to find out which one is the best. In this sense, we need a way to select a small set of subtrees for consideration and guarantee that the pruned tree is optimal.

Minimal cost-complexity pruning

A way to compare the various pruned trees is to estimate each subtree's misclassification rate and select the one with a lower misclassification rate. We call the resubstitution estimate r(t) to the proportion of training error in a node, and R(T) the resubstitution estimate for the overall tree ($\tilde{T}$) so that

$$r(t) = 1 - \max P(j/t)$$

where $\max P(j/t)$ is the probability of the majority class in node t based on the training data, and

$$R(T) = \sum_{t \epsilon \tilde{T}} R(t) = \sum_{t \epsilon \tilde{T}} r(t).p(t)$$

where *p(t)* is the proportion of the sample in that node.

Then, one way to select a pruned tree would be to choose the one with the lowest resubstitution rate for the overall tree ($\tilde{T}$). However, R(T) is not a good measure because it decreases as more nodes are created, favoring always bigger trees that lead to overfitting problems. This can be solved by introducing a penalisation parameter, so that the resubstitution estimate increases as the tree grows further. This resubtitution rate with the additional penalisation parameter is what is known as the cost-complexity measure, $R_\alpha(T)$, and it is given by

$$R_\alpha(T) = R(T) + \alpha|T|$$

Where |T| is the number of terminal nodes and $\alpha \geq 0$ would be a real number called the complexity parameter. The complexity parameter controls the trade-off between the subtree's complexity and its fit. If $\alpha=0$, then the largest tree would be selected. Whilst if $\alpha \longrightarrow \infty$, then the tree size would equal 1, or what is the same, a single root node will be selected.

Therefore, the objective is to find a value of α that properly penalises the overfitting and gives the tree the right complexity. This is achieved when $R_\alpha(T)$ is minimised.

$$R_\alpha\big(T(\alpha)\big) = \min_{T \leq T_{max}} R_\alpha(T)$$

Since there are at most a finite number of subtrees of T (tree with maximum number of possible nodes), $R_\alpha(T(\alpha^i))$ yields different values for only finitely number of α's. $T(\alpha^i)$ continues to be the minimising tree when α increases until a jump point is reached. At this point, a new tree $T(\alpha^{i+1})$, that is smaller than $T(\alpha^i)$ becomes the new minimising tree. Every new jumping point for α can be found solving the following inequality:

$$\alpha < \frac{R(t) - R(T_t)}{\tilde{T} - 1}$$

And this resulting α is also known as the weakest link cutting.

In summary, based on the learning sample we would find a sequence of optimal pruned subtrees such that $T_0 > T_1 > T_2 \ldots > T_k$, where T_k has only the root node T_0, and every subtree minimises $T(\alpha^i)$. The algorithm starts with $\alpha^0=0$, and then find the optimal subtree for every weakest link cutting. The output after the algorithm would be: a sequence of optimal trees such as $T_0 > T_1 > T_2 \ldots > T_k$; and a sequence of weakest link cuts such as $\alpha^0 < \alpha^1 < \ldots < \alpha^k$.

To finally obtain a classifier, all that remains to be done is selecting the one αi with the lowest misclassification rate for future predictions. It can be obtained using a simple cross-validation test, which implies calculating out-of-sample errors for every optimal tree and

selecting the one with the lowest out-of-sample error, also known in the machine learning literature as test error.

Appendix 2. Independent variables description

1. **Dividend Yield, DY:** MSCI EMU *I/B/E/S* [8]dividend yield.
2. **Price to Book Value, PBV:** EMU DS price to book value.
3. **Price to trailing earnings, PET:** It is the ratio between *I/B/E/S* MSCI EMU index and the trailing (past 12 month) *I/B/E/S* earnings per share for the *I/B/E/S* MSCI EMU.
4. **Price to forward earnings, PEF:** It is the ratio between *I/B/E/S* MSCI EMU index and the forward (expected next 12 month) *I/B/E/S* earnings per share for the *I/B/E/S* MSCI EMU.
5. **Price to 5Y average trailing earnings, PE5T:** It is the ratio between *I/B/E/S* MSCI EMU index and the 5 year average of trailing *I/B/E/S* earnings per share for the *I/B/E/S* MSCI EMU.
6. **Price to 5Y average forward earnings, PE5F:** It is the ratio between *I/B/E/S* MSCI EMU index and the 5 year average of forward *I/B/E/S* earnings per share for the *I/B/E/S* MSCI EMU.
7. **Earnings momentum, EMOM:** It is the ratio between *I/B/E/S* MSCI EMU analyst company's upgrade revisions minus analyst company's downgrades and the total company's analyst revisions.
8. **Euribor 3 months, EUR3M.**
9. **Swap 2Y, SWAP2:** 2 years Euro Swap rate.
10. **Swap 10Y, SWAP10:** 10 years Euro Swap rate.
11. **Term Spread 3M-10Y, Slope310:** It is the difference between the 10 years swap rate and the Euribor 3M rate.
12. **Term Spread 2Y-10Y, Slope210:** It is the difference between the 10 years swap rate and the 2 years swap rate.

[8] Institutional Brokers' Estimate System (I/B/E/S) is a service from Refinitiv that gathers and compiles stock data and analyst estimates.

13. **Nymex oil price level, OIL.**

14. **Inflation, INF:** EU Harmonized CPI Y/Y change (NSA).

15. **VIX index, VIX.**

16. **ASW IG Corporates, IG:** Asset swap spread of EMU Investment Grade index (Bank of America Merrill Lynch Euro Investment Grade index).

17. **ASW HY Corporates, HY:** Asset swap spread of EMU High Yield index (Bank of America Merrill Lynch Euro High Yield index).

18. **Default Spread, Spread:** It is the difference between HY spread and IG spread.

19. **TED Spread, TED:** It is the difference between the Euribor 3 months and ECB main refinancing rate.

20. **EU Consumer confidence, CONSCONF.**

21. **EU Business Climate Indicator, BUSCLIM.**

22. **EU Industrial Confidence, INDCONF.**

23. **EU Retail Confidence, RETCON.**

24. **OECD EU Leading Indicator, OECDLEAD.**

25. **Moving averages rules, MOVAVGS:** These rules give buy and sell signals depending on the short and long moving averages of prices. If the short moving average is above the long average, then there is a buy signal, and if it is below, there is a sell signal. We analyse monthly MA rules with short MA with t= 2 months, and long MA with t=6.

26. **On-balance volume rules, OBV:** It measures buying and selling pressure as a cumulative indicator that adds volume on up days and subtracts volume on down days.

$$OBV_t = Cummulative\ volume_{t-1} + volume_t$$

Appendix 3. Out-of-sample results

Table A3: Out-of-sample forecasting results including all simple linear regressions

Variable	MSFE	ROS	t-stat	p-value
ΔIND_CONF	0,0007	0,65	-4,88	0.00***
ΔSWAP_2Y	0,0008	0,60	-4,01	0.00***
ΔCONS_CONF	0,0009	0,59	-4,84	0.00***
ΔRET_CONF	0,0010	0,53	-4,83	0.00***
ΔMov.Avg	0,0018	0,14	-5,04	0.00***
ΔOil	0,0019	0,11	-5,60	0.00***
ΔBUS_CLIM	0,0020	0,06	-4,90	0.00***
ΔVIX	0,0020	0,04	-4,67	0.00***
EMOM	0,0021	0,03	-4,70	0.00***
TED_Spread	0,0021	0,02	-1,38	0.08**
ΔEMOM	0,0021	0,02	-2,96	0.00***
Oil	0,0021	0,01	-1,39	0.08**
OBV	0,0021	0,01	-2,20	0.02***
HY	0,0021	0,01	-0,28	0,39
ΔINF	0,0021	0,01	-6,16	0.00***
PE5T	0,0021	0,00	-0,09	0,47
ΔOBV	0,0021	0,00	-3,07	0.00***
Spread_HY.IG	0,0021	0,00	-0,01	0,50
VIX	0,0021	0,00	-0,03	0,49
ΔSpread_HY.IG	0,0021	0,00	-1,63	0.05**
ΔHY	0,0021	0,00	-1,23	0,11
ΔIG	0,0021	0,00	-0,31	0,38
Benchmark	*0,0021*	*0,00*	*n.d*	*n.d*
ΔTED_Spread	0,0021	0,00	2,55	0,99
ΔSlope_3m.10Y	0,0021	0,00	2,31	0,99
PE5F	0,0022	-0,01	0,14	0,56
BUS_CLIM	0,0022	-0,01	2,39	0,99
ΔDY	0,0022	-0,01	2,86	1,00
PET	0,0022	-0,01	0,21	0,58
ΔPET	0,0022	-0,02	3,18	1,00
IG	0,0022	-0,02	2,12	0,98
ΔSlope_2Y.10Y	0,0022	-0,03	4,28	1,00
PEF	0,0022	-0,03	0,35	0,63
ΔPEF	0,0022	-0,04	4,55	1,00
RET_CONF	0,0022	-0,05	4,62	1,00
Pruned_Tree	0,0023	-0,06	0,91	0,82
ΔPBV	0,0023	-0,06	3,98	1,00
ΔPE5T	0,0023	-0,07	4,54	1,00
ΔPE5F	0,0023	-0,07	4,55	1,00
Random Forest	0,0023	-0,07	2,00	0,98
Bagging	0,0023	-0,10	1,67	0,95
IND_CONF	0,0024	-0,14	5,51	1,00
DY	0,0025	-0,18	2,80	1,00
INF	0,0025	-0,18	1,61	0,95
PBV	0,0027	-0,25	1,85	0,97
Slope_3m.10Y	0,0027	-0,25	3,38	1,00
CONS_CONF	0,0028	-0,31	5,34	1,00
Slope_2Y.10Y	0,0028	-0,31	3,54	1,00
Boosting	0,0030	-0,40	3,84	1,00
Tree	0,0032	-0,49	3,05	1,00
ΔEUR_3M	0,0036	-0,69	4,83	1,00
ΔSWAP_10Y	0,0044	-1,05	3,89	1,00
SWAP_10Y	0,0061	-1,88	5,15	1,00
SWAP_2Y	0,0063	-1,96	5,01	1,00
EUR_3M	0,0066	-2,08	4,92	1,00
Mov.Avg	0,0153	-6,17	14,15	1,00
dOECDLEAD	0,0608	-27,55	4,29	1,00
ΔOECDLEAD	0,0774	-35,30	27,54	1,00

, and *** indicate significance at the 10%, 5% and 1% levels, respectively.*
MSFE is the mean squared forecast error for the out-of-sample period.
ROS is the Cambell and Thompson R^2 statistic.
n.d. means "no data".
t-stat and p-value are the statistic and the p-value of the Diebold-Mariano test for predictive accuracy. Univariate linear regressions are calculated as: $r_{t+1} = \alpha_t + \beta_t * x_{i,t}$, where α_t and β_t are the univariate estimations up to time t, $x_{i,t}$ is the predictor i at period t, and r_{t+1} is the equity risk premium at t+1.

Chapter 3

Macro determinants of non-performing loans: a comparative panel analysis between consumer and mortgage loans

3.1 Introduction

Understanding non-performing loans (NPLs) and which factors explain their evolution is an essential task at a microeconomic and macroeconomic level. On the one hand, risk departments of financial institutions must comply with regulators, calculate non-performing loans for all their credit portfolios, and report them regularly. Maintaining a low level of non-performing loans is a key performance indicator that directly impacts the bank's profitability and capital levels. On the other hand, non-performing loans are also crucial in the macroeconomic sphere because aggregated non-performing loans are good indicators of the banking system's credit quality. Thus, a rapid rise in non-performing loans in the system could become a systemic problem that might interrupt the normal flow of credit and threaten financial stability and economic growth.

The economic literature has provided extensive evidence that the business cycle and asset prices shocks are among the principal systemic factors behind non-performing loans changes. Shifts in macroeconomic conditions or financial shocks can affect borrowers' economic conditions and increase their propensity to default. Manz (2019) presents a systematic literature review on the determinants of non-performing loans covering papers published from 1987 to 2017. Between these studies, Keeton and Morris (1987) find that local economic conditions combined with the low performance of various economic sectors are responsible for differences in loan losses recorded by different banks. Jimenez and Saurina (2006) show that past credit growth and other common macroeconomic determinants have a significant and lagged impact on Spanish banks delinquencies. Louzis et al. (2012) find that variables such as real GDP growth rates, unemployment rate, lending rates and public debt have a substantial effect on the level of non-performing loans. And Nkusu (2011) analysed the linkage between non-performing loans and the macroeconomic performance of 26 advanced economies from 1998 to 2009, revealing that poor macroeconomic performance could be associated with increasing non-performing loans in developed economies.

Although the literature covering the primary macroeconomic determinant of non-performing loans is extensive, data availability is a major constraint, and there is no standardised approach to study causal relationships and forecasting performance. On the one hand, if one attends to the nature of the data reviewed, NPLs research can be divided into two groups. The first one analyses specific bank data to highlight the role of macroeconomic factors on credit quality. The second group works with aggregated data produced by central banks,

financial supervisors, or any other national or international office. On the other hand, if one attends to the econometric techniques used in the literature, approaches implemented are multiple, but the most used ones are panel data techniques. This article works with aggregated data at the country level and using panel models.

The benefits of using panel data procedures are multiple, so they are frequently used in this type of research. As Hsiao (1986) emphasises, panel data benefits from more extensive data sets, with more variability and less collinearity. Besides, panels can control the individual heterogeneity of the cross-sections included in the analysis without actually observing it. According to Pesaran (2015), the literature on panel data could be divided into three broad categories depending on their assumptions about the relative number of cross-sectional units (N) and the number of time periods (T). First, "large N, small T" literature, also known as micro panels. Second, "large N, large T" literature, also known as macro panels. And finally, "small N, large T" literature.

Panel category is an important feature because, depending on the sample size, different estimation approaches can be undertaken to calculate the impact of macroeconomic drivers on non-performing rates. So far, most of the research covering this subject has focused its attention either on micro panels or macro panels, and little attention has been paid to intermediate situations where T is large, and N is small. An important exception in the literature is in Rinaldi and Sanchis-Arellano (2006), which focus their research on an unbalanced panel, including seven euro area countries over a period spanning from 1989Q3 to 2004Q2.

Our paper examines the influence of several macroeconomic factors on non-performing loans and contributes to the empirical literature in several ways. First, adding research to the scarce "large T, small N" panel literature on the economic determinants of NPLs. Often, practitioners and regulators have to research and forecast over intermediate-sized unbalanced panels because obtaining and aggregating the data can be costly or impossible to achieve. Not every country has the same data quality and availability. In particular, this paper arranged a dataset with quarterly data that spans from 1992Q1 to 2019Q4 for a sample of 8 developed and emerging economies: United States, Spain, Mexico, Turkey, Colombia, Peru, Argentina and Chile. We selected this heterogeneous and short sample of countries as a real example of geographies regularly analysed by risk analysts or researchers of one of Spain's largest commercial banks. Second, this study examines whether macroeconomic factors impact differently on NPLs across loan categories. Louzis et al. (2012) proved that macroeconomic

and bank-specific variables influence Greek NPLs categories differently. We expand this analysis to a broader set of countries and for two loan portfolios, consumer loans and mortgages. Third, it contributes to the debate of whether to pool or not to pool the data when the panel is unbalanced, heterogenous, and not all the cross-sections are long enough to undertake individual time series. Fourth, it adds out-of-sample forecasting exercises to complement and confirm in-sample findings. Fifth and last, it enriches the debate on the possibility of using Fixed Effects-Within Groups (FE-WG) models in dynamic panels with a long number of time observations.

The rest of the paper is organised as follows. Section 3.2 provides an overview of the theoretical and empirical literature on the macroeconomic determinants of non-performing loans. Section 3.3 explains the data and the model structure selected. We discuss essential time series characteristics such as unit roots or cointegration, the existence of cross-section dependence, or the heterogeneity of the estimated parameters. Section 3.4 shows the estimation results. In section 3.4.1. in-sample estimations are calculated throughout five econometric models: Pooled OLS, Fixed Effects-Within Groups (FE-WG), 2SLS instrumental variables, 3SLS instrumental variables and Dynamic Mean Groups (DMG) estimations. In section 3.4.2. we present the out-of-sample forecast of all the estimated models. To conclude, section 3.5 summarises our main results.

3.2 Literature review

There is abundant literature examining the influence of the macroeconomic environment on credit quality, being this quality measured as non-performing loans (NPLs), probability of default (PD), the loss given default (LGD), loan loss provisions (LLP) or any other measure of credit risk. One of the first papers that linked economic activity with loan losses was Keeton and Morris (1987). In this paper, the authors examined a sample of 2,470 insured commercial banks in the United States from 1978-1985. They found that local economic conditions, combined with the low performance of various economic sectors, were responsible for differences in loan losses recorded by different banks. After this initial paper, the interest in the evolution of the NPLs and their macroeconomic determinants has dramatically increased.

This study arranges the literature covering NPLs and their determinants into two groups. The first one looks at specific bank data to highlight the role of macroeconomic and other

particular determinants. In this research category, Salas and Saurina (2002) take a panel of Spanish commercial and savings banks during 1985-1997, compare the determinants of problematic loans, and conclude that macroeconomic conditions matter for loans performance. Jiménez and Saurina (2006) combine macroeconomic and microeconomic variables to explain the aggregate NPLs of Spanish Commercial and Savings Banks from 1984 to 2002. The paper aggregates individual banks' information and finds solid empirical support of a positive, although quite lagged, relationship between rapid credit growth and loan losses. It also finds that other common macroeconomic determinants such as GDP growth or real rates have a significant impact. Fofack (2005) examines individual banks from a group of Sub-Saharan countries. Using unbalanced panel data of 16 countries with a total of 90 observations for the period 1993-2003, results highlight a strong causality between non-performing loans and economic growth, real exchange rates, real interest rates, net margins and interbank loans. Quagliariello (2007) analyses an unbalanced panel of 207 Italian banks over almost two decades to understand whether delinquencies behave cyclically in Italy. Espinoza and Prasad (2010) examine the effect of various macroeconomic and banking variables in the NPLs ratio in the Gulf Cooperation Council (GCC) countries. Looking at 80 banks level data for the period 1998-2008, the paper finds that macroeconomic variables such as non-oil real GDP growth, stock market returns, interest rates, world trade growth and the VIX index, along with other various banking variables, determine the level of NPLs ratio in Gulf countries. Glen and Mondragón-Vélez (2011) study the effects of the business cycle on commercial bank loan portfolios' performance across major developing economies. Using panel data methods, they find that economic growth is the primary driver of loan portfolios performance and that interest rates have second-order effects. Louzis et al. (2012) use dynamic panel data methods to examine non-performing loans' determinants across three different loan categories (commercial loans, consumer loans and mortgages) in the Greek Sector. Results show that variables such as real GDP growth rates, unemployment rate, lending rates and public debt have a substantial effect on the level of non-performing loans. Klein (2013) investigates the non-performing loans in Central, Eastern and South-Eastern countries (CESEE) in 1998-2011. Panel data analysis concludes that NPL rates respond to macroeconomic variables such as GDP growth rates, unemployment rates, exchange rates, inflation or global risk aversion indicators (VIX).Similarly, Messai and Jouini (2013) examine a panel of 85 banks in Italy, Spain and Greece for 2004-2008 and find economic growth and bank profitability to reduce NPLs, while unemployment, real interest rates, and low credit quality positively influence them. Also,

Mohaddes et al. (2017) examine whether a tipping point exists for real GDP growth in Italy, above which the non-performing loans rate falls significantly. The paper uses a dynamic panel-threshold model and finds that real GDP growth above 1.2% is associated with a significant decline in NPL ratios if sustained for several years. Finally, Ghosh (2015,2017) examine the state-level banking industry and regional economic determinants of NPLs for commercial and saving institutions across 50 US States and districts of Columbia for 1984-2013. These articles find that higher state real GDP and real personal income growth rates, and higher state housing prices reduce NPLs.

The second group of papers, to which our analysis pertains, explains and predicts NPLs looking at aggregated financial system ratios. For instance, Brookes et al. (1994) model mortgage arrears of building societies in the United Kingdom and highlight the role of rising inflation in creating mortgages defaults. Rinaldi and Sanchis-Arellano (2006) investigate the linkage between household non-performing loans ratio and various macroeconomic variables in the Eurozone area using a panel of seven euro area countries. The paper finds that the set of macroeconomic variables included in the model contribute to explain a good portion of the variations of loan arrears. These macroeconomic variables include GDP growth rates, ratios of indebtedness to income, inflation, lending rates and financial and housing wealth. Nkusu (2011) analyses with panel regressions the links between NPL and macroeconomic performance on a sample of 26 advanced economies and confirm that adverse macroeconomic developments are associated with rising NPL. Similarly, De Bock and Demyanets (2012) turns to dynamic panel regressions to determine the factors driving banks asset quality in 25 emerging markets during 1996-2010. Beck et al. (2013) study the macroeconomic determinants of NPL across 75 countries during the past decade. The paper finds applying static and dynamic panels that real GDP growth, share prices, exchange rates and lending interest rates explain NPL rates.

All of the above studies work with either micro or macro panels. However, it is difficult to find in the literature studies using large T and short N panels. Relevant exceptions are Rinaldi and Sanchis-Arellano (2006) and Skarica (2014). The first article focuses its research on an unbalanced panel, including seven euro area countries (N=7) over a period from 1989Q3 to 2004Q2. The paper proposes a dynamic error correction model (ECM) that captures both short-run and long-run effects on NPLs. On the other hand, Skarica (2014) examines quarterly data from 2007 to 2012 for seven Central and Eastern European countries and, using fixed effects

estimators, find both unemployment and inflation rates to increase the growth of NPLs while real GDP growth has adverse effects.

This study, similarly to Rinaldi and Sanchis-Arellano (2006) and Skarica (2014), focuses its analysis on "large T, small N" unbalanced panels, but considering a different dataset of countries and periods, and breaking down households loans into consumer and mortgages loans.

3.3 The data and the model

3.3.1 Data

This study covers quarterly data for a sample of 8 developed and emerging economies, namely Spain (ESP), Mexico (MXC), the United States (USA), Turkey (TUR), Colombia (COL), Peru (PER), Chile (CHL) and Argentina (ARG). The countries' sample is heterogeneous, and the number of cross-sections is scarce, but we select it for two reasons. First, we would like to contribute to the "large T, small N" panel literature that analyses credit risk's macroeconomic determinants. Often, obtaining the disaggregated NPL data can be costly and not always possible. Hence practitioners have to deal with short datasets that cannot be classified as micro or macro panels. The second reason is that we select this sample of countries because they represent a true example panel of countries that a Spanish commercial bank analyses under IFRS9[1] macroprudential policy.

Concerning the factors included in the models, we select the ratio of non-performing loans (NPLs) as the endogenous variable and as a measure of credit quality. We calculate NPLs at the macroeconomic level from the consolidated balance sheets of each country's banking sector, and these are broken down into consumer and mortgages portfolios. As Louzis et al. (2011) notice, macroeconomic variables may impact each type of NPLs in different ways because the business cycle does not always affect in the same manner each kind of loan and collateral.

[1] IFRS9 is an International Financial Reporting Standard which addresses the accounting for financial instruments. Under this Standard, financial entities are required to update their economic forecasts on regular basis, and how these macroeconomic forecasts impact expected credit losses (ECL). In this line, financial entities elaborate macroeconomic predictions for all the countries in which they have credit businesses, and decide which are the best econometric techniques to produce those predictions.

NPLs rates are calculated as the ratio between those loans that have been in arrears for some time and the total amount of loans in the portfolio. These cross-country NPLs should be interpreted with caution because they not always reflect the same reality across geographies[2]. Countries use different accounting standards and prudential frameworks, which could result in differences in the volumes and timing of credit loss provisions and difficult cross-countries comparisons. Appendix 1 describes the key variables used in modelling NPLs rates and details the data sources.

As explanatory variables, we consider six factors. One is the lagged endogenous variable (*npl_lag*) because delinquency rates tend to show a high degree of persistence. And five[3] macroeconomic factors that the economic literature (see, for a review, Manz (2019)) have proven to impact delinquency rates significantly.

The first selected factor is the past credit growth rate (*cred*). We include this input as a possible measure of leverage. Jiménez and Saurina (2006) produced clear evidence of lagged relationships between the credit cycle and credit risk. They showed that a rapid increase in loan portfolios is positively associated with an increase in non-performing loan ratios later on.

The second factor is real GDP (*gdp*). This variable is selected as a measure of the state of the economy. GDP dynamics are closely related to households' capacity to meet their obligations, as the expansionary phases of the economy tend to be associated with higher incomes to service debts and lower NPL levels.

As a third factor, we choose short term real rates (*rates*). Real rates measure the cost of servicing debt, and we should expect a positive relationship between delinquencies and short term real rates. Higher real rates could translate into a higher debt burden and make it more challenging for borrowers to repair their debts. Nonetheless, there are economic situations where this positive relationship could not work well[4].

[2] Baudino et al. (2018) identify and measure non-performing assets across a wide range of countries and regions, and find that differences in calculations can be significant and materially impact provisions needed.

[3] We included only five macroeconomic explanatory variables to maintain our model as straightforward as possible while gathering all relevant information. Our panels cross-section dimension is small (N=8), and including more economic factors could create collinearity problems. Besides, for several factors there was no reliable data for all the countries (e.g. liquid long term sovereign yields), or the inclusion of these variables for some countries seemed less relevant (eg. Inflation or exchange rates for the United States and Spain).

[4] Imagine situations where central banks react quickly to a falling economic activity lowering reference rates significantly. In these situations, though debt burden would be alleviated, delinquency rates could rise at fast pace.

Finally, the fourth and fifth explanatory variables are real home prices indices (*hpr*) and equity indices (*equity*). These two variables gather wealth effects. A downturn in financial and/or real asset prices leads to more defaults via wealth effects and a decline in the collateral value.

Table 1 summarises the six explanatory variables defined above and their expected impact in line with the economic literature. The table also informs the number of observations for each country in the panel, ranging from March 1992 to December 2019. Furthermore, Appendix 2 represents all the series selected in this study.

Table 1: Explanatory variables, expected impact on NPL and observations per country

Symbol	**Explanatory Variable**	**Expected Sign**	
npl_lag	Lagged dependent variable	(+)	
cred	Lagged Credit Volumes	(+)	
gdp	Real GDP	(-)	
rates	Real Short Term Rates	(+)/(indet)	
hpr	Real Home Prices	(-)	
equity	Equity Index	(-)	
Country	**Observations**	**Starting date**	**End date**
ARG	57	mar.-06	dic.-19
CHL	45	mar.-09	dic.-19
COL	62	dic.-04	dic.-19
ESP	82	dic.-99	dic.-19
MEX	57	mar.-06	dic.-19
PER	71	sep.-02	dic.-19
TUR	37	mar.-11	dic.-19
USA	113	mar.-92	dic.-19

Summarising, we have two unbalanced panels (consumer and mortgages loans) with a T=37-113 and N=8, where the most extended time series belongs to the United States, which starts in March 92 and ends on December 19, and the shortest to Turkey, which begins on March 11 and ends on December 19. The size of the panel is a relevant feature because, as Pesaran (2015) notes, the choice of an appropriate estimator depends on the relative size of the number of cross-sections (N) and the number of periods (T). A large body of the panel literature (see, for instance, Baltagi et al. (1995) or Rinaldi and Sanchis Arellano (2006)) recommend a seemingly unrelated regression (SURE) approach due to Zellner (1962) for situations with large T and small N, and where N is reasonably small relative to T. This paper will follow literature recommendations and calculate, along with other panel data methods, a mix between SURE procedures and instrumental variables. We will focus our attention on the 2SLS and 3SLS techniques, which use instrumental variables and correct cross-section heteroskedasticity and contemporary correlation in the errors across the equations, such as SURE methods.

Indeed, the 3SLS estimation due to Zellner and Theil (1962) could be considered a combination of SURE techniques with instrumental variables.

3.3.2 The model

Following the discussion in the previous section, the empirical analysis is based on the following expression:

$$npl_{n,t}^{j} = f\left(npl_{n,t-1}^{j}, cred_{n,t}^{j}, gdp_{n,t}, rates_{n,t}, hpr_{n,t}, equity_{n,t}\right) \quad (1)$$

Where $npl_{n,t}^{j}$ is the log of non-performing loans rate for country n, at time t and for portfolio j, and it proxies the credit risk of that portfolio (either consumer or mortgage loans). The NPLs rate is calculated as the ratio of the number of non-performing loans in the portfolio to the total amount of outstanding loans in that same portfolio. $cred_{n,t}^{j}$ is the log of the level of credit of portfolio j, for country n, at time t; $gdp_{n,t}$ is the log of real GDP for country n at time t; $rates_{n,t}$ is short term real rates for country n at time t; hpr is the log of real house prices for country n at time t, and $equity_{n,t}$ is the log of equity indices for country n at time t.

Due to many temporal observations and the cross-section nature of datasets, time-series procedures also become relevant. Therefore, before applying any econometric model, we need to check three relevant features. First, to test for the presence of unit roots and possible cointegration relationships. To avoid spurious estimations, we must confirm that dependent and independent variables do not have unit roots and, if it is the case, whether there are stable long-term relationships between the variables. Second, the presence of cross-section dependence across each country must be tested. This cross-section dependence can take two primary forms. Either it depends on the relative position of the units in the space, spatial correlation, or all the individual units are subject to the same set of common global factors. If cross-section dependence is significant and not dealt with, standard panel estimators can result in misleading inference and even inconsistent estimators (Pesaran, 2015). Third, the poolability of the data must be tested. Long panels allow to estimate separate regressions for each cross-unit, and therefore it is natural to question the assumption of parameters homogeneity ($\beta_n = \beta \ \forall n$), also called the pooling assumption. Rejecting this assumption means that traditional procedures for the estimation of pooled methods can produce inconsistent and potentially

misleading estimates of the parameters in our dynamic panel models unless the slope coefficients were, in fact, homogeneous.

3.3.2.1 Unit Roots and Cointegration

Since Levin and Lin (1992, 1993) seminal work, the study of unit roots and cointegration has played an increasing role in analysing panel data, especially for panel data sets with a large number of temporal observations. Our panel contains a time dimension that is relatively large, and thereby stationarity and cointegration checks become relevant to obtain consistent estimators.

Regarding panel unit roots tests, these can be divided into two groups. The first one, known as the "first generation unit root tests", assumes no relationships between the different cross units. And "second-generation unit roots tests" take into account potential residual cross-section dependence in panels. This study runs six first-generation tests: Levin-Lin-Chu, Breitung, Hadri, Im-Pesaran-Shin, Fisher-ADF and Fisher-PP. And one from the second generation, the CIPS test proposed in Pesaran (2007).

Appendix 3 shows the results obtained for the unit root panel tests. Outcomes differ depending on the tests used, the presence of constants and trends, or the number of lags included in the analysis. In general, most of the tests, and particularly the Fisher[5] stationarity tests, support the presence of unit roots in almost all the variables studied. Only interest rates seem to reject the existence of unit roots often.

Stationarity tests also confirm that series with unit roots are integrated of order one. Instead of using quarterly differences, we use annual differences[6] to check the order of integration because it might have more economic sense in this type of study. Thus, Rinaldi and Sanchis-Arellano (2006) argue that taking annual differences makes more sense since the NPL rate variability with respect to the previous quarter might be small and because households do not update their decisions very often.

Having assessed that most of the variables are integrated of order one except for real short-term rates, the next step is to check the existence of long-run equilibrium. That is, searching

[5] Some literature mention that these tests would be best suited to check stationarity in unbalanced panels. See for example Stata 13 manual, xtunitroot section. https://www.stata.com/manuals13/xtxtunitro pdf ot.

[6] For example, the annual difference of the Real GDP (Δgdp_t) would be calculated as $\ln(gdp_t) - \ln(gdp_{t-4})$.

for the existence of possible cointegration relationships among the selected variables. To do so, we use the Pedroni test. Pedroni (1999,2004) extend Engle and Granger (1987) methodology to test cointegration in panel data structures. Pedroni proposes seven statistics that allow for heterogeneous intercepts and trend coefficients across cross-sections. Consider the following regression:

$$y_{i,t} = \alpha_i + \lambda_{i,t}t + \beta_i z_{i,t} + e_{i,t} \quad (2)$$

Where α_i and λ_i are individual and trend effects, which may be set to zero if desired.

Pedroni tests check whether $e_{i,t}$ is I(0) or not by running the following auxiliary residual regression:

$$e_{i,t} = \rho_i e_{t-1} + u_{i,t} \quad \forall \text{ cross-section} \quad (3)$$

Under the null hypothesis of no cointegration ($\rho_i = 0$), seven tests are built. Four of them check for the homogeneous alternative hypothesis of $\rho_i = \rho < 1$, and these are known as panel statistics tests or within-dimension tests. While the remaining three statistics check for the heterogeneous alternative hypothesis of $\rho_i < 1$, and these tests are referred to as the group statistics of the between-dimension statistics.

Table 2: Pedroni Cointegration Tests

Consumer Loans

	No Ctant & No Trend		Ctant & No Trend		Ctant & Trend	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Panel v-Statistic	-2,11	0,982	1,13	0,130	1,47	0,071
Panel rho-Statistic	2,76	0,997	0,66	0,746	0,73	0,769
Panel PP-Statistic	1,72	0,957	-0,67	0,250	-0,72	0,235
Panel ADF-Statistic	2,22	0,987	0,20	0,580	-0,04	0,484
Group rho-Statistic	0,15	0,561	0,17	0,568	0,94	0,826
Group PP-Statistic	-2,31	0,011	-3,05	0,001	-2,74	0,003
Group ADF-Statistic	-1,96	0,025	-0,43	0,334	-1,03	0,151

Series: npl_cons,cred_cons(-4),gdp,hpr,rates,equity.
Pedroni Cointegration Tests (Ho=No cointegration).

Mortgage Loans

	No Ctant & No Trend		Ctant & No Trend		Ctant & Trend	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Panel v-Statistic	-1,62	0,947	0,88	0,189	1,10	0,136
Panel rho-Statistic	2,58	0,995	0,99	0,840	1,61	0,946
Panel PP-Statistic	2,42	0,992	0,68	0,751	1,05	0,854
Panel ADF-Statistic	2,78	0,997	0,20	0,579	-1,11	0,134
Group rho-Statistic	2,47	0,993	2,22	0,987	2,15	0,984
Group PP-Statistic	1,45	0,927	1,23	0,890	0,39	0,654
Group ADF-Statistic	1,16	0,876	0,14	0,556	-1,97	0,025

Series: npl_mtg,cred_mtg(-4),gdp,hpr,rates,equity.
Pedroni Cointegration Tests (Ho=No cointegration).

Table 2 shows the results for the seven tests proposed by Pedroni, allowing for heterogenous intercepts and trends. We check the existence of cointegrating relationships between NPLs and the rest of the explanatory variables, and most of the tests do not reject the null hypothesis of no cointegration. In the case of consumer loans, only two tests reject the null hypothesis of no cointegration at a 5% confidence level when there is no constant and no trend. In the case of mortgages, only one test accepts the hypothesis of cointegration when there is a constant and trend. Therefore, we conclude that there is sufficient evidence to discard the existence of a stable long-run relationship for arrears, both in consumer loans and mortgages.

Our results contradict those obtained in Rinaldi and Sanchis-Arellano (2006). They find cointegrating relationships and select an error correction model to calculate the relationships between the macroeconomic factors and non-performing loans rates. In this study, since we could not find such relationships between the variables, we select a simple[7] autoregressive distributed lag structure, also known as an ARDL model, where we take first annual differences for all the variables that were integrated of order one. In this structure, we include the first lag

[7] We select the simplest lag structure in this article, because after running an exercise comparing different lag structures between ARDL(1,0) and ARDL(2,2), we found that estimation results obtained were not significantly different. Hence for simplicity, we chose the simplest autoregressive structure that was possible.

of the endogenous variable, *gdp*, *rates,* and *equity* are contemporaneous to the dependent variable, and *credit* appears lagged four quarters. We introduce past credit growth lagged one year because Salas and Saurina (2002) and Jimenez and Saurina (2006) proved that one-year past credit growth has a positive and significant impact on loan losses.

Hence, after the stationarity and cointegration tests, equation 1 takes the following form:

$$\Delta npl_{n,t}^{j} = \gamma_n \Delta npl_{n,t-1}^{j} + \beta_{1,n}\Delta cred_{n,t-4}^{j} + \beta_{2,n}\Delta gdp_{n,t} + \beta_{3,n}\Delta hpr_{n,t} + \beta_{4,n} rates_{n,t} + \beta_{5,n}\Delta equity_{n,t} + \eta_n + \varepsilon_{n,t} \quad (4)$$

Where the superscript *j* denotes the portfolio (either mortgages or consumer loans), *n* is the country, *t* the quarter, η_n are fixed effects for each country, and $\varepsilon_{n,t}$ are the residuals.

3.3.2.2 Cross Section Dependence test

When the panel's time-series dimension is large, cross-section dependence becomes a relevant issue because correlation across the different units might arise. This relation between the cross-section units can appear due to two sources of dependence. The first one is known as spatial dependence or local spillovers. It occurs when a cross-section unit's value at a point in space is related to the value at nearby cross-units. It can be explained by distance, and nearby units show higher relationships than distant ones. For instance, as the price for homes rises in a neighbourhood, people tend to move to adjacent areas, increasing costs there too. The second type of dependence is known as global dependence, and it appears when unobserved common factors affect each unit to the extent that it does not depend on the distance. An example of this type of dependence would be a global shock such as a worldwide pandemia or a global financial shock.

Ignoring cross-section dependence can have severe consequences. Conventional panel estimators such as fixed or random effects, which assume cross-sectionally independent errors, can result in misleading inference, inefficient estimators and biased standard errors, and even inconsistent estimators if unobserved common factors correlate with regressors.

To test cross-section dependence, we implement the cross-section test developed by Pesaran (2004), which has good properties in samples of practically any size. It is robust to

various settings, and it is suitable for unbalanced panels. We implement the test in three steps. First, we estimate by ordinary least squares (OLS) each cross-section separately.

$$y_{n,t} = \alpha_n + \beta_n' x_{n,t} + u_{n,t} \tag{5}$$

Where $\alpha_n$and β_n for n=1,2,…, N, are assumed to be fixed unknown coefficients, and $x_{n,t}$ is a k-dimensional vector of regressors.

Second, we compute the pair-wise correlations of the residuals obtained.

$$\hat{\rho}_{nm} = \hat{\rho}_{mn} = \frac{\sum_{t=1}^{T} \hat{u}_{n,t}\, \hat{u}_{m,t}}{\left(\sum_{t=1}^{T} \hat{u}_{n,t}^2\right)^{1/2} \left(\sum_{t=1}^{T} \hat{u}_{m,t}^2\right)^{1/2}}$$

Where $\hat{\rho}_{nm}$ are the pair-wise correlations of the residuals of each regression. And third, we run the Pesaran (2004) test suitable for unbalanced panels, where the null hypothesis assumes no cross-sectional dependence in model errors ($\hat{\rho}_{nm}$=$\hat{\rho}_{mn} = 0\ for\ n \neq m$).

$$CD = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{n=1}^{N-1} \sum_{m=n+1}^{N} \sqrt{T_{nm}}\, \hat{\rho}_{nm} \right)$$

Table 3: Pesaran (2004) CD modified test for cross-sectional dependence in panels

Pesaran CD Test		
Portfolio	z-stat	p-value
Consumer Loans	**5.16*** **	0,000
Mortgages	-0,45	0,649

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3 shows that Pesaran CD tests reject the null hypothesis of no cross-sectional dependence for consumer loans portfolios, but not for mortgages, where we accept the null hypothesis. Results are striking because they suggest that consumer loans portfolios might be more affected by global trends than mortgages. A possible explanation that could partly explain this difference lies in the economic factors that affect both portfolios. Consumer credit evolution could be driven by global factors such as economic growth, real disposable income, or the cost of funding. Simultaneously, mortgages could be affected by these same global

factors and other local ones such as real estate valuations, which are also explained by local factors such as demographics or specific public policies. In any case, results suggest that cross-section dependence is present in one of the portfolios, and cross-sectionally independent errors cannot be assumed in panel estimations.

3.3.2.3 Testing for Poolability

Another relevant question that arises when dealing with panels with long T is whether we should pool the data or not. Long panels allow to estimate separate regressions for each cross-unit, and therefore it is natural to question the assumption of parameters homogeneity ($\beta_n = \beta \ \forall n$), also called the pooling assumption.

Baltagi et al. (2008) explain that for panel data studies with large N and small T, it is usual to pool the observations, assuming the slope coefficients' homogeneity. However, with the increasing dimension of panel data sets, some researchers, including Robertson and Symons (1992), Pesaran and Smith (1995), and Pesaran et al. (1999) have questioned the poolability of the data across heterogeneous cross units. Instead, they argue in favour of heterogenous estimates that can be combined to obtain homogeneous estimates. The decision "to pool or not to pool" spans vast literature, and there are arguments in favour and against it. However, as Croissant and Millo (2019) note, it could be safely stated that data-rich environments favour heterogeneous estimates, while the appeal of pooling restrictions becomes higher the smaller the dataset.

Our dataset's intermediate nature makes us wonder whether our panels should keep the pool hypothesis or not. While there are countries such as the United States or Spain with periods that cover 108 or 77 quarterly observations (more than 20 years), others like Turkey or Chile only have 32 and 40 quarterly observations, respectively. These later periods might seem large within the panel data literature. However, they only cover between 7 and 9 years of data, which, given the nature of our data, does not seem long enough to rely on these individual country regressions' consistency properties.

To see how the pooling assumption fits our dataset, we will estimate each country individually again using equation (4) structure and observe how those estimations vary across countries and line up with the economic literature. Table 4 displays ordinary least squares (OLS) estimates for each country and portfolio

Table 4: OLS estimations for each country

Consumer Loans. Individual OLS estimations

Country	Intercept	Δnpl_cons (t-1)	Δcred_cons (t-4)	Δgdp	Δhpr	real rates	Δequity
arg	-0,22***	0,65***	0,72***	-0,93***	-0,09	-0,10	-0,00
chl	-0,12**	-0,11**	1,39***	-2,02	0,16	0,08	0,08
col	-0,06**	0,8***	0,56***	0,31	-0,52**	-1,70	-0,05
esp	-0,01	0,84***	0,56***	0,33	-0,73***	0,05	-0,06
mxc	-0,08***	0,22*	1,13***	-0,74**	-1,00	0,10	-0,3***
per	-0,01	0,78***	0,18**	-0,62	0,06	0,43	-0,03
tur	-0,05	0,82***	0,55***	-1,01***	0,47*	1,16**	0,01*
us	-0,00	0,82***	0,20	-0,20	-0,09	0,60**	-0,08

Mortgages Loans. Individual OLS estimations

Country	Intercept	Δnpl_cons (t-1)	Δcred_cons (t-4)	Δgdp	Δhpr	real rates	Δequity
arg	-0,14**	0,51***	1,15***	-1,05**	2,25**	-0,32	-0,06
chl	0,19***	0,12***	-0,94*	-0,26	-1,93***	-4,75***	-0,20***
col	0,03	0,80***	0,05	-1,76***	0,16**	1,12	-0,08
esp	-0,08	0,94***	0,59***	2,81	-1,17**	-2,24	-0,09
mxc	0,01	0,59***	0,27**	-2,05***	-0,09	-0,37	0,01
per	0,00	0,93***	0,06	-0,53	0,25**	0,90	-0,02
tur	0,044	0,82***	0,28	-2,23***	-0,07	0,93*	0,02
us	-0,01	0,86***	0,37***	-0,11	-0,48**	0,22	-0,01

****p < 0.01; **p < 0.05; *p < 0.1, with Newey & West heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators.*

Table 4 shows that the variability between the coefficients across countries is broad and might point to heterogeneous estimations. However, some of the estimates' signs are surprising and suggest that maybe the time dimension in several countries is not long enough to undertake individual regressions. Thus, for example, the impact of GDP on delinquencies is not always negative. Δhpr alternates signs from one country to another, and similar behaviour can be found in real rates. Even the persistency expected in delinquency rates is not always found. Therefore, estimates obtained in Table 4 indicate that individual results do not seem very consistent and that the time dimension in several units could be not long enough to eliminate the poolability assumption.

3.4 Estimating results

3.4.1 In-sample estimation

Previous section outcomes indicate that we have a "large T, small N" unbalanced dynamic panel; with a T significantly larger than N, for which we were not able to find cointegration relationships; and where cross-section dependence is present in the case of consumer loans panel, but not for mortgages. Furthermore, several cross-units small-time dimensions have made us impose the poolability assumption. Hence, considering all these aspects, we estimate the following dynamic panel.

$$\Delta npl_{n,t}^{j} = \gamma \Delta npl_{n,t-1}^{j} + \beta_1 \Delta cred_{n,t-4}^{j} + \beta_2 \Delta gdp_{n,t} + \beta_3 \Delta hpr_{n,t} + \beta_4 rates_{n,t} + \beta_5 \Delta equity_{n,t} + \eta_n + \varepsilon_{n,t} \quad (6)$$

As in equation (4), the superscript j denotes the portfolio (either consumer loans or mortatges), n is the country, t the quarter, η_n are the unobservable individual-specific effects which are invariant over time t, Δ indicates annual differences of the natural logarithms of the variables, and $\varepsilon_{n,t}$ are the residuals.

To estimate equation (6), we use five estimation methods. The first two, the Poolability (Pool) and the Fixed Effects Within Group (FE-WG) models, estimate parameters with ordinary least squares. These are well known traditional panel data models (e.g. Baltagi (2005), Wooldridge (2010) or Pesaran (2015)) frequently applied when panel equations are static, but which show bias and inconsistencies[8] when equations are dynamic, or errors are not independently and identically distributed. Despite these problems, we estimate them because they are easy to calculate and will allow us to compare coefficients variability between the different models. To remove the endogeneity problem generated by the lagged dependent variable, we also calculate two instrumental variables methods: the two-stage least-squares (2SLS) and the three-stage least-squares (3SLS). Both methodologies remove endogeneity problems but not always result in efficient estimators. The 2SLS could not give efficient estimations when there exist contemporaneous cross-section correlations on the residuals. In

[8] Baltagi (1998) explains that when the lagged dependent variable is among the regressors, OLS estimators would be biased and inconsistent since the lagged component $\Delta npln_{n,t-1}$ would be a function of the unobservable individual effects η_n. For the fixed effects (FE) estimator, the within transformation removes η_n, but $\Delta npln_{n,t-1}$ would be still correlated with $\varepsilon_{n,t}$ and estimators would be biased (Nickell (1981)). In fact, the within estimator will be biased of O(1/T) and its consistency will depend upon T being large.

such cases, the 3SLS proposed by Zellner and Theil (1962) would provide consistent and efficient estimators since it considers the correlation among the errors of each of the simultaneous equations of interest. Moreover, the 3SLS also adjusts the weighting matrix for potential heteroskedasticity problems by estimating the coefficients within a Generalized Least Square (GLS) framework. Therefore, the 3SLS technique appears as the most efficient one among all described to calculate equation (6), particularly in consumer loans, where cross-section dependence exists in the residuals as seen in previous sections.

Additionally, to compare several techniques, we also calculate Dynamic Mean Group estimations. This technique was introduced by Pesaran and Smith (1995), and they showed that the average estimator of the heterogenous parameters of each cross-section could lead to consistent[9] estimations as long as N and T tend to infinity.

Table 5 displays estimated coefficients and their p-values for consumer loans and mortgages. Overall, the estimated models went in line with past literature. Almost every explanatory variable included in the models show the expected signs and are statistically significant. Moreover, results also confirm heterogeneity in the macro sensibilities between consumer and mortgages loans portfolio. Thus, for instance, the impact of the GDP changes in delinquencies is more substantial in mortgages than in consumer loans, and it can be said the same in the case of real house prices or real interest rates. This result should not be surprising since fundamental loans characteristics such as the cash-flow structure or the collateral behind the loans could be very different.

Focusing the attention on the explanatory variables, estimations of the lagged dependents confirm the persistent nature of delinquency rates in both portfolios. Besides, coefficients seem similar for all models. Mortgages exhibit slightly higher persistence than consumer loans, but in both portfolios, the coefficients are high and significant, and none of them near or above one.

[9] The Dynamic Mean Group (DMG) estimation is appropriate when both T and N (macropanels) are sufficiently large. In this article, the number of cross-countries is small, and averaging countries estimates of a such a small sample could not lead to consistent estimators, specially if the individual estimations diplay high variability. We included this model to also have a macropanel example, and to check how its estimates would compare to the rest.

Table 5: Panel data estimations for consumer loans and mortgages

Panel Estimations											
Consumer Loans						Mortgages					
	POOL	FE_WG	IV-2SLS	IV-3SLS	DMG		POOL	FE_WG	IV-2SLS	IV-3SLS	DMG
(Intercept)	-0.01				**-0.06****	(Intercept)	0.00				0.00
	(0.832)				(0.015)		(0.467)				(0.961)
Δnpl_cons(t-1)	**0.70*****	**0.67*****	**0.64*****	**0.64*****	**0.60*****	Δnpl_mtge(t-1)	**0.74*****	**0.72*****	**0.69*****	**0.70*****	**0.67*****
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Δcred_cons(t-4)	**0.27*****	**0.44*****	**0.46*****	**0.44*****	**0.66*****	Δcred_mtge(t-4)	**0.36*****	**0.39*****	**0.41*****	**0.41*****	0.26
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.212)
Δgdp	**-0.84*****	**-0.88*****	**-0.99*****	**-0.91*****	**-0.69****	Δgdp	**-1.20*****	**-1.22*****	**-1.31*****	**-1.22*****	-0.76
	(0.007)	(0.005)	(0.002)	(0.002)	(0.024)		(0.000)	(0.002)	(0.000)	(0.000)	(0.210)
Δhpr	-0.05	-0,11	-0.11	**-0.12****	-0.29	Δhpr	**-0.23*****	-0.21	-0.24	**-0.26*****	0.08
	(0.594)	(0.398)	(0.435)	(0.047)	(0.127)		(0.007)	(0.370)	(0.326)	(0.002)	(0.875)
rates	**0.34****	0.08	0.12	0.13	0,14	rates	0.19	**0.45*****	**0.47*****	**0.49****	-0,81
	(0.010)	(0.679)	(0.579)	(0.313)	(0.457)		(0.190)	(0.002)	(0.001)	(0.039)	(0.220)
Δequity	**-0.07*****	**-0.06*****	**-0.06*****	**-0.06*****	**-0.06***	Δequity	**-0.09*****	**-0.09*****	**-0.10*****	**-0.09*****	**-0.05****
	(0.000)	(0.000)	(0.000)	(0.000)	(0.087)		(0.000)	(0.007)	(0.003)	(0.000)	(0.018)
Num. obs.	492	492	492	492	492	Num. obs.	492	492	492	492	492

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

t-stats and p-values calculated with clustered standard errors in POOL, FE-WG and IV-2SLS, but not in IV-3SLS and DMG.

R^2 fot DMG is calculated as the multiple-R^2.

Instruments used in 2SLS and 3SLS methods for Δnpl_cons(t-1) are: Δnpl_cons(t-2),Δnpl_cons(t-3) and Δnpl_cons(t-4).

Other explanatory variables which are statistically significant and show the expected sign in all models are the past credit growth *(Δcred_cons(t-4)* and *(Δcred_mtge(t-4)*), the real GDP growth (Δ*gdp*), and equity markets growth (Δ*equity*). The former shows a positive effect on delinquency rates in every estimation method, confirming that rapid credit growth in previous periods leads to higher credit risk in future periods, as suggested in Jiménez and Saurina (2006). Outcomes also demonstrate that real GDP growth and increases in equity markets harm NPLs. Only the DMG would reject the GDP's significance in mortgages loans and shows a slightly higher elasticity. Finally, all models find a small and negative relationship between non-performing rates and growth in equity markets. Coefficients are near zero and suggest that the effects of equity markets on arrears are not large.

The remaining two explanatory variables, real interest rates and growth in real house prices, show more mixed evidence. Overall, both exhibit expected signs, positive in the case of real interest rates, pointing to a positive impact of higher funding costs on delinquencies and negative in the case of real house prices. In the case of house prices, most models conclude that this variable is not significant, whilst in the case of real rates, it does not seem a statistically significant variable, but it does for mortgages.

Finally, another exciting outcome that shows Table 5 is the coefficients obtained through the different econometric techniques. It can be observed that the variability of the coefficients is low, especially between the FE-WG and Instrumental variables models. A plausible

explanation could be that the number of temporal observations of the panels is high enough to remove a great deal of the bias the FE-WG estimations might show when T is not large enough. This thought is also confirmed when we look at the lagged dependent variables' coefficient estimations. Instrumental variables and FE-WG seem very similar, suggesting that FE-WG bias might be partially removed. At the same time, Pooled OLS exhibit higher coefficients pointing to an upward bias of these estimators.

In summary, in-sample results obtained are in line with the literature, macroeconomics factors are essential explanatory drivers of delinquency rates, and the heterogeneity found in the two portfolios' coefficients makes it advisable to estimate the portfolios separately. Furthermore, the unbalanced nature of the panel (where some countries have a large number of time observations to carry out individual time series, but others do not) made us impose the poolability assumption. Nonetheless, at the same time, T seems large enough to give very similar results between the FE-WG estimations and instrumental variables, suggesting that the FE-WG bias in dynamic estimations could be partially removed.

3.4.2 Out-of-sample estimation

As a final robustness check, we run a simple out-of-sample forecasting exercise to check whether the macroeconomic factors selected can improve non-performing forecasts compared to a naïve benchmark model. We choose the individual autoregressive model as a benchmark because it is simple and seems more appropriate than a historical mean average due to the proven persistence in delinquency growth rates.

To produce out-of-sample predictions, we divide the sample data into two sub-samples. The first one, known as the estimation sample, over which we obtain our panel parameters estimations. And a second one, known as the forecasting sample, over which we realise predictions. For each of the models, we forecast dynamically three forecasting periods. A very short term period which covers four quarters (one year), a short period which covers eight quarters (two years), and a medium-term period which covers twenty quarters (five years). Hence the out-of-sample panel regression for each model and country is given by:

$$\widehat{\Delta npl}^{j}_{n,t+i} = \hat{\gamma}\Delta npl^{j}_{n,t+i-1} + \hat{\beta}_1\Delta cred^{j}_{n,t+i-4} + \hat{\beta}_2\Delta pib_{n,t+i} + \hat{\beta}_3\Delta hpr_{n,t+i} + \hat{\beta}_4 rates_{n,t+i} + \hat{\beta}_5\Delta equity_{n,t+i} + \hat{\eta}_n \quad (7)$$

Where $\hat{\gamma}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$ *and* $\hat{\eta}_n$ are estimations obtained through the different panel methods.

To evaluate forecast accuracy, we select the Mean Squared Forecast Error (MSFE) metric. We compare forecasts in terms of the relative MSFE ratio, which calculates the MSFE of the panel chosen models out of the MSFE of the benchmark for each one of the three forecasting periods selected.

$$Relative\ MSFE = \frac{MSFE_{benchm}}{MSFE_{model}}$$

If the relative MSFE is greater than 1, the panel model forecasts will outperform those obtained by the benchmark, while if the relative MSFE is lower than 1, then the opposite would be true.

Table 6: Out-of-sample forecasting results. Relative MSFE

Relative MSFE: Consumer Loans

	Pool				FE-WG				IV-2SLS				IV-3SLS				DMG			
	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge
Argentina	1,0	**3,5**	**1,5**	**2,0**	**1,9**	**3,6**	**1,6**	**2,4**	**1,9**	**4,3**	**1,7**	**2,6**	**1,8**	**3,7**	**2,1**	**2,5**	**0,5**	**1,4**	**1,0**	**1,0**
Chile	**5,2**	**3,7**	**3,0**	**4,0**	**1,3**	0,5	0,4	0,7	**1,5**	0,6	0,5	0,9	**1,4**	0,7	**1,1**	**1,1**	0,6	0,5	0,5	0,6
Colombia	0,5	**2,3**	1,0	**1,3**	**1,3**	**4,7**	**1,1**	**2,4**	**1,1**	**4,9**	**1,1**	**2,4**	**1,2**	**6,3**	**1,3**	**2,9**	**2,0**	**10,4**	**0,6**	**4,4**
Mexico	0,3	0,4	**1,1**	0,6	**1,6**	**2,0**	**2,5**	**2,0**	**1,3**	**1,7**	**2,6**	**1,9**	**1,5**	**2,1**	**2,5**	**2,0**	**0,6**	**2,6**	**1,5**	**1,6**
Peru	**4,2**	**1,4**	**1,2**	**2,2**	**2,9**	**2,7**	**1,6**	**2,4**	**3,3**	**2,6**	**1,6**	**2,5**	**3,3**	**2,8**	**1,7**	**2,6**	**3,3**	**4,0**	**0,9**	**2,8**
Spain	**2,2**	**2,5**	**3,3**	**2,7**	0,3	0,2	**1,1**	0,5	0,3	0,2	**1,1**	0,5	0,3	0,2	**1,8**	0,8	**5,1**	**1,1**	**2,1**	**2,8**
Turkey	0,6	0,5	-	0,5	0,7	0,5	-	0,6	0,6	0,4	-	0,5	0,8	0,8	-	0,8	0,7	0,9	-	0,8
United States	**1,0**	0,0	**1,3**	0,8	**9,2**	**0,1**	**2,2**	**3,8**	**5,9**	0,1	**2,2**	**2,7**	**5,9**	0,1	**3,4**	**3,1**	**0,1**	0,0	0,3	0,1

Relative MSFE: Mortagages

	Pool				FE-WG				IV-2SLS				IV-3SLS				DMG			
	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge	1Y Fcast	2Y Fcast	5Y Fcast	Avge
Argentina	**12,9**	**2,1**	**1,9**	**5,6**	**14,1**	**2,5**	**2,1**	**6,2**	**11,5**	**2,5**	**2,2**	**5,4**	**14,9**	**7,9**	**5,6**	**9,5**	**5,1**	**2,2**	**2,1**	**3,1**
Chile	**2,5**	**2,9**	**2,4**	**2,6**	**2,0**	0,3	0,3	0,9	**2,2**	0,3	0,3	0,9	**2,3**	0,4	1,0	**1,2**	**2,2**	**3,8**	**1,2**	**2,4**
Colombia	0,2	**6,9**	**1,6**	**2,9**	0,2	**11,0**	**1,4**	**4,2**	0,1	**11,6**	**1,4**	**4,4**	0,2	**12,0**	**2,1**	**4,8**	**0,2**	**13,2**	**1,9**	**5,1**
Mexico	**1,4**	0,9	**1,5**	**1,3**	**2,5**	**1,5**	**1,9**	**2,0**	**2,5**	**1,5**	**2,0**	**2,0**	**2,9**	**1,5**	**4,3**	**2,9**	0,7	**0,3**	**1,4**	0,8
Peru	**3,1**	**8,7**	**2,5**	**4,8**	**1,5**	**10,6**	**2,2**	**4,8**	**1,5**	**9,9**	**2,1**	**4,5**	**2,0**	**8,7**	**3,7**	**4,8**	**5,6**	**3,0**	**2,5**	**3,7**
Spain	**2,0**	**2,7**	**4,4**	**3,0**	0,6	**1,3**	**2,1**	**1,4**	0,5	1,3	2,1	**1,3**	0,0	0,1	0,1	0,0	0,5	**1,6**	**3,7**	**1,9**
Turkey	**2,3**	**5,6**	-	**3,9**	**1,1**	**2,9**	-	**2,0**	**1,0**	**3,1**	-	**2,1**	**1,3**	**7,9**	-	**4,6**	0,4	**1,1**	-	0,7
United States	0,1	0,5	0,2	0,3	0,0	0,4	0,1	0,2	0,0	0,3	0,1	0,2	0,0	0,5	0,3	0,3	0,1	0,6	0,1	0,2

Relative MSFE > 1, model beats bechmark; Relative MSFE < 1, model underperforms bechmark.
Relative MSFE = $MSFE_{benchmark}$ /$MSFE_{model.}$
5Y forecast for Turkey are not included because sample data for this country includes only 7 years.
Benchmark estimations obtained from an AR(1) structure such as: $\Delta npl_{n,t} = \Delta npl_{n,t-1} + \varepsilon_{n,t}$ where n represents the country and t the quarter selected.

Table 6 exhibits out-of-sample forecast results for each country and the forecasting period. In general, it could be said that panel models tend to beat the benchmark in most of the forecasting windows, and economic factors show forecasting ability. Comparing the different econometric techniques, all the models display similar results, and there are no significant differences among forecasting windows, ruling out in this sample that models have greater or

lesser predictive capacity depending on the interval of time that is being predicted. Maybe, 3SLS relative MSFE shows slightly better results at an aggregated level in both portfolios than the rest of the methodologies.

At the portfolio level, forecast performance is better in mortgages than in consumer loans. In the later portfolio, there are three countries, Chile, Spain and Turkey, where panels cannot beat the benchmark. Only in Spain's case, this situation changes for the 5Y forecast period, suggesting that economic factors could help further in the medium-term horizons than in shorter ones in this country. In mortgages loans, panel models outperform the benchmark in all countries except the United States, where forecasting techniques do not perform well in any forecasting window.

In summary, and for the sample studied, it could be stated that there are more cases in which models beat the benchmark than not. Hence, it could be said that economic factors play an essential role in forecasting non-performing loans, in line with the outcomes obtained in-sample.

3.5 Conclusions

This study explores the macroeconomic determinants of non-performing loans (NPLs) in two international unbalanced dynamic panels, one for consumer loans and another for mortgages. In particular, it explores a dataset with quarterly data that spans from 1992 to 2019 for a sample of 8 developed and emerging economies: United States, Spain, Mexico, Turkey, Colombia, Peru, Argentina and Chile. This short sample of countries was selected for two reasons: the first one is because obtaining disaggregated data for NPLs is difficult and costly since not all countries produce these numbers. The second one is that it aims to contribute to the scarce "large N, Small T" panel literature covering the macro determinants of NPLs rates. This type of datasets problem is usually found by practitioners when estimating and forecasting delinquency rates for macroprudential reasons.

Due to the panels' sample structure, we checked several time-series features such as unit roots and cointegration relationships, the existence of cross-sectional dependence, and whether pooling the estimation coefficients made sense in these samples. Results obtained indicate that series do not show cointegration relationships, consumer loans residuals have cross-section dependence but mortgages residuals don't, and the scarce data in some countries recommend

maintaining the poolability assumption. Hence, we built autoregressive distributed lag models taking annual first differences, imposed parameters homogeneity ($\beta_n = \beta \ \forall n$) and calculated instrumental variable models that correct the endogeneity problems created by the lagged dependent variable and consider the possibility of contemporary cross-section correlation. Specifically, we estimated instrumental variables through two-stage least squares (IV-2SLS) and three-stage least squares (IV-3SLS). Besides, we complemented estimations with two commonly used models in the static panel literature, Pooled OLS and Fixed Effects Within Groups (FE-WG), and a macro panel technique, the Dynamic Mean Groups (DMG). We added these three techniques to compare coefficients across models.

In-sample estimations obtained were in line with the economic literature. Delinquency rates show a persistent nature, past credit growth has a positive impact on NPLs, and an acceleration in real GDP, real house prices and equity markets, along with lower funding costs, lead to lower NPLs. Estimations also confirmed that consumer loans and mortgages display very similar sensibility to the economic cycles. Yet, the elasticities are slightly different and invite to estimate these two portfolios independently. In addition, in-sample estimations also showed that coefficients obtained through FE-WG and instrumental variables look very similar, making us wonder if the sample T is long enough to remove FE-WG bias in these dynamic panels. Finally, the out-of-sample forecast confirmed that most models tend to beat the benchmark and economic factors play an essential role in forecasting non-performing loans.

This chapter' results could be of relevance for academics and practitioners in several ways. First, this study contributes to the extensive literature exploring the macroeconomic determinants of non-performing loans and finds similar results, but focusing on "large T, small N" panel techniques where literature is not abundant. Second, contrary to other research, we could not find cointegration relationships between NPLs rates and the macroeconomic factors selected. Third, the paper contributes to the debate of whether to pool or not pool in unbalanced panels, with a significant time dimension but a concise number of cross-sections. In this situation, FE-WG estimations could remove their bias and produce consistent estimators. Finally, the article includes out-of-sample forecasts as a robustness check for some of the results obtained in-sample.

Looking ahead, future research to complement and improve the present study should research further the lag structure of the autoregressive distributed panel models and investigate whether other explanatory variables, not only the dependent one, are endogenous due to reverse

causality. In this line, estimating a PVAR structure could offer a robustness check of the panel regression results. Moreover, including richer time series data and further cross-section observations disaggregating time series by categories would help obtain more efficient estimators and improve credit risks forecasts. In this sense, this paper should encourage authorities and regulators to produce more extended NPLs time series as disaggregated and broad as possible.

References

Álvarez-Román, L., and García-Posada Gómez, M., 2019. Modelling Regional Housing Prices in Spain. Banco de España. Documentos de Trabajo Nº 1941.

Baltagi, B., 1998. Panel Data Methods. A note prepared for the *Handbook of Applied Economic Statistics*, edited by Aman Ullah and David E.A.Giles, Marcel Dekker, New York.

Baltagi, B., 2005. *Econometric Analysis of Panel Data*, J. Wiley & Sons.

Baltagi, B., Bresson, G., and Pirotte, A., 2008. To pool or not to pool? In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Springer-Verlag.

Baltagi, B., Griffin, J.M. , and Rich, D.P., 1995. Airline deregulation: the cost pieces of the puzzle. *International Economic Review*, *36, 245-259*.

Baudino, P., Orlandi, J., and Zamil, R., 2018. The identification and measurement of non-performing assets: a cross-country comparison. Bank for International Settlements. *FSI Insights on policy implementation Nº 7.*

Beck, R., Jakubik, P., and Piloiu, A., 2013. Non-performing Loans. What Matters in Addition to the Economic Cycle. *Working Paper Series Nº 1515.* European Central Bank.

Berger, AN., DeYoung, R. 1997. Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance 21. 849–870*.

Brookes, M., Dicks M., and Pradhan, M., 1994. An Empirical Model of Mortgage Arrears and Repossessions. *Economic Modelling*, *11, 134–144*.

Croissant, Y., and Millo, G., 2019. *Panel Data Econometrics*, R. John Wiley & Sons Ltd.

De Bock, R. and Demyanets, MA., 2012. Bank asset quality in emerging markets: determinants and spillovers. *IMF working paper WP/12/71*, International Monetary Fund, 2012.

Dimitrios, A., Helen, L., and Mike, T., 2016. Determinants of non-performing loans: evidence from Euroarea countries. *Finance Research Letters, 18. 116–119*.

Engle, R.F., and Granger, C.W.J., 1987. Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, *55, 251-276*.

Espinoza, R., and Prasad, A., 2010. Non-performing Loans in the GCC Banking System and their Macroeconomic Effects. *IMF Working Paper 10/224*.

Fernández de Lis, S., Pagés, J.M., and Saurina, J., 2000. Credit Growth, Problem Loans and Credit Risk Provisioning in Spain. *Banco de España Working Paper 18.*

Fofack, H., 2005. Non-performing Loans in Sub-Saharan Africa: Causal Analysis and Macroeconomic Implications. *World Bank Policy Research Working Paper Nº 3769*.

Ghosh, A., 2015. Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. *Journal of Financial Stability, 20*. *93–104*.

Ghosh, A., 2017. Sector-specific analysis of non-performing loans in the US banking system and their macroeconomic impact. *Journal of Economics and Business 93*. *29–45*.

Glen, JD., and Mondragón-Vélez, C., 2011. Business Cycle Effects on Commercial Bank Loan Portfolio Performance in Developing Economies. *Review of Development Finance, 1, 150-165*.

Hsiao, C., 1986. Analysis of Panel Data, Cambridge University Press, Cambridge, MA.

Jiménez, G. and Saurina, J., 2006. Credit Cycles, Credit Risk, and Prudential Regulation. *International Journal of Central Banking, 2(2), 65-98*.

Keeton, WR., and Morris, C.S., 1987. Why do Banks' Loans Losses Differ? *Economic Review*, *Federal Reserve Bank of Kansas City, 3-21*.

Klein, N., 2013. Non-performing Loans in the CESEE. Determinants and Macroeconomic Performance. *IMF Working Paper 13/72.*

Levin, A. and Lin, C.F., 1992. Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties. Discussion Paper, Department of Economics, University of California at San Diego.

Levin, A. and Lin, C.F., 1993. Unit root tests in panel data: New results. Discussion Paper, Department of Economics, University of California at San Diego.

Louzis, D.P., Vouldis, A.T., and Metaxas, V.L., 2012. Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking and Finance*, *36 (4), 1012-1027*.

Makri, V., Tsagkanos, A., and Belles, A., 2014. Determinants of non-performing loans: the case of eurozone. *Panoeconomicus 61:193–206.*

Manz, F., 2019. Determinants of non-performing loans: what do we know? A systematic review and avenues for future research. *Management Review Quarterly, 69. 351-389.*

Messai, A., and Jouini, F., 2013. Micro and macro determinants of non-performing loans. *International Journal of Economics and Financial Issues*, *3. 852-860*

Mohaddes, K., Raissi, M., and Weber, A., 2017. Can Italy Grow Out of Its NPL Overhang? A Panel Threshold Analysis. *Economics Letters*, *159, 185-189.*

Nickell, S.J., 1981. Biases in dynamic models with fixed effects, *Econometrica, 49, 1417-1426.*

Nkusu, M., 2011. Non-performing Loans and Macrofinancial Vulnerabilities in Advanced Economies. *IMF Working Paper 11/161.*

Pedroni, P., 1999. Critical values for cointegration tests in heterogeneous panels with multiple regressors', *Oxford Bulletin of Economics and Statistics, 61, 653– 670.*

Pedroni, P., 2004. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory, 3, 579– 625.*

Pesaran, M.H., 2004. General diagnostic tests for cross-section dependence in panels. *CESifo Working Papers N° 1233.*

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22, 265-312.

Pesaran, M.H., 2015. Time Series and Panel Data Econometrics. Oxford University Press.

Pesaran, M.H., and Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics, 68, 79-113.*

Quagliariello, M., 2007. Banks' riskiness over the business cycle: a panel analysis on Italian intermediaries. *Applied Financial Economics, 17:119–138.*

Rinaldi, L., and Sanchis-Arellano, A., 2006. Household Debt Sustainability, What Explains Household Non-Performing Loans? An Empirical Analysis. *European Central Bank Working Paper Series 570.*

Robertson, D. and Symons, J., 1992. Some strange properties of panel data estimators, *Journal of Applied Econometrics, 7, 175-189*.

Salas, V., and Saurina, J., 2002. Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks. *Journal of Financial Services Research, 22(3), 203-224.*

Skarica, B., 2014. Determinants of non-performing loans in Central and Eastern European Countries. *Financial Theory and Practice, 38*. 37-59.

Smith, R. , and Fuertes, A.M., 2016. Panel Time Series, Birkbeck College, London.

Wooldridge, J., 2010. Econometric Analysis of Cross-Section and Panel Data. MIT press. 2nd edition.

Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregations. *Journal of the American Statistical Association Journal*, 57*, Nº.298, 348-368.*

Zellner, A. and Theil, H., 1962. Three-Stage Least Squares: Simultaneous Estimations of Simultaneous Equations. *Econometrica, 30 (1), 54-78.*

Appendix 1. Non-performing loans data sources

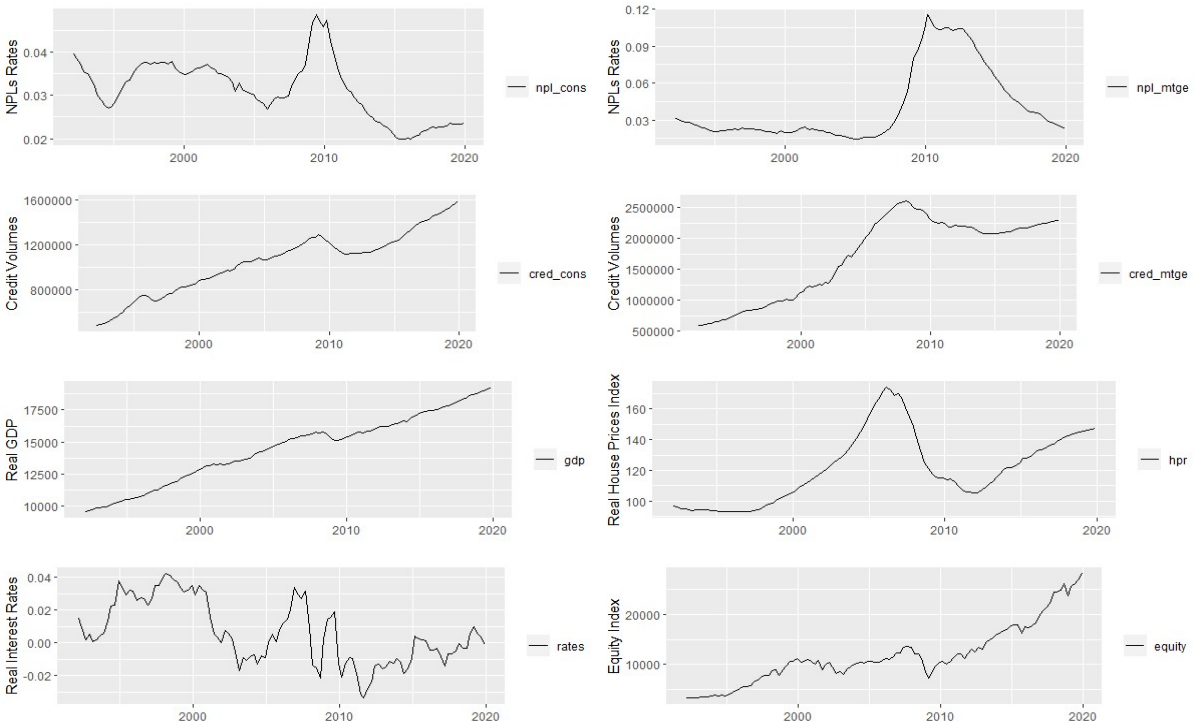
Table A1: Non-performing loans data sources

Country	Porftfolio	Variable	Source
USA	Consumer loans	Loan Delinquency Rate: Consumer Loans: All Insured Comml Banks (EOP, SA,%)	Federal Reserve Board / Haver Analytics
	Mortgages	Loan Delinquency Rate: Residential Real Estate Loans: All Comml Banks (EOP,SA,%)	Federal Reserve Board / Haver Analytics
Spain	Consumer loans	Dudosos, Bienes de consumo TOTAL/Crédito para adquisición de bienes de consumo	https://www.bde.es/webbde/es/estadis/infoest/series/be0413.xlsx
	Mortgages	Dudosos, Adquisición y rehabilitación de vivienda/Crédito para adquisición y rehabilitación de viviendas	https://www.bde.es/webbde/es/estadis/infoest/series/be0413.xlsx
Mexico	Consumer loans	Commercial Bank: Nonperforming Loans: Total Consumer Loans (% of Total)	Banxico / Haver Analytics
	Mortgages	Commercial Bank: Nonperforming Loans: Housing Loans (% of Total)	Banxico / Haver Analytics
Turkey	Consumer loans	Non-performing Consumer Loans (banking system)/Consumer Loans (Banking Sector)	https://www.bddk.org.tr/BultenAylik/en/Home/Gelismis
	Mortgages	Non-performing Housing Loans (banking system)/Housing Loans (Banking Sector)	https://www.bddk.org.tr/BultenAylik/en/Home/Gelismis
Colombia	Consumer loans	Consumo: Indicador de Calidad Tradicional %	https://www.superfinanciera.gov.co/publicacion/60950
	Mortgages	Vivienda: Indicador de Calidad Tradicional %	https://www.superfinanciera.gov.co/publicacion/60950
Perú	Consumer loans	Morosidad Crédito Consumo	http://www.sbs.gob.pe/app/stats_net/stats/EstadisticaBoletinEstadistico.aspx?p=1#
	Mortgages	Morosidad Crédito Hipotecario	http://www.sbs.gob.pe/app/stats_net/stats/EstadisticaBoletinEstadistico.aspx?p=1#
Argentina	Consumer loans	Households Consumer: Non-performing financing / Total financing (%)	http://www.bcra.gov.ar/PublicacionesEstadisticas/Informe_mensual_sobre_bancos.asp
	Mortgages	Households - Mortgage and pledge-backed: Non-performing financing / Total financing (%)	http://www.bcra.gov.ar/PublicacionesEstadisticas/Informe_mensual_sobre_bancos.asp
Chile	Consumer loans	Consumo: Cartera con morosidad de 90 días o más (Consolidada)	http://www.cmfchile.cl/portal/estadisticas/606/w3-propertyvalue-28914.html
	Mortgages	Vivienda: Cartera con morosidad de 90 días o más (Consolidada)	http://www.cmfchile.cl/portal/estadisticas/606/w3-propertyvalue-28914.html

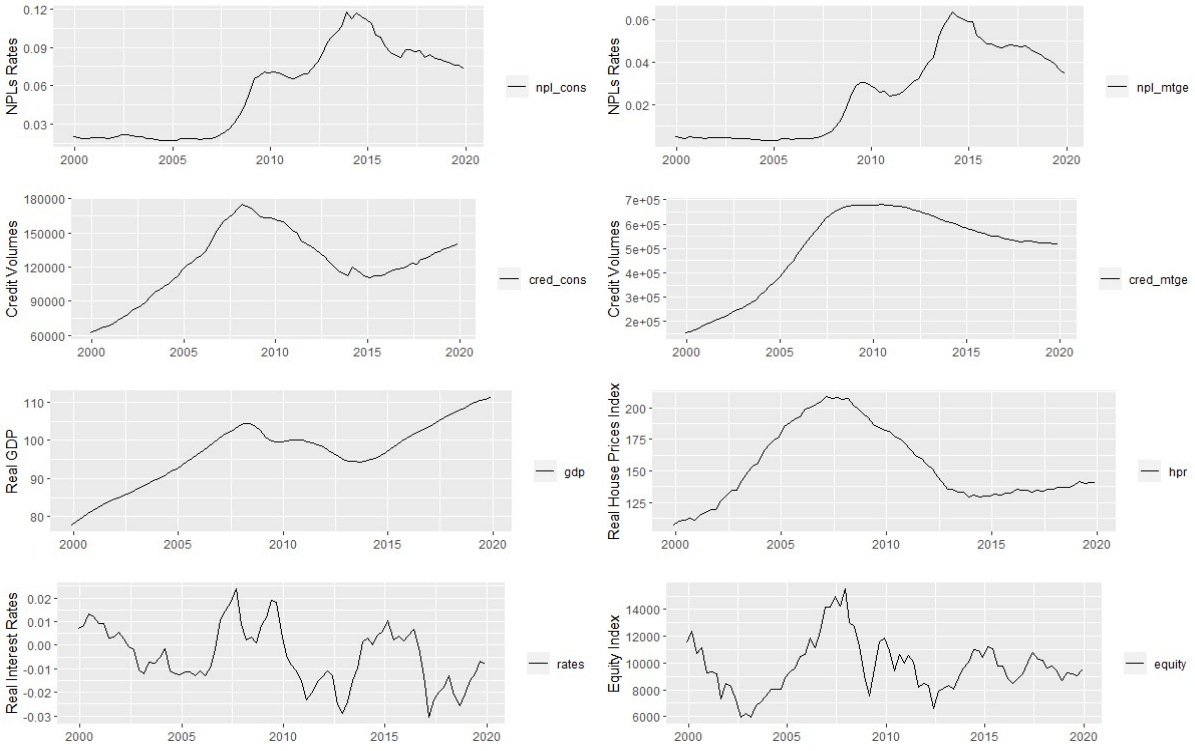
Appendix 2. Country variables charts

Figure A2: Country variables charts

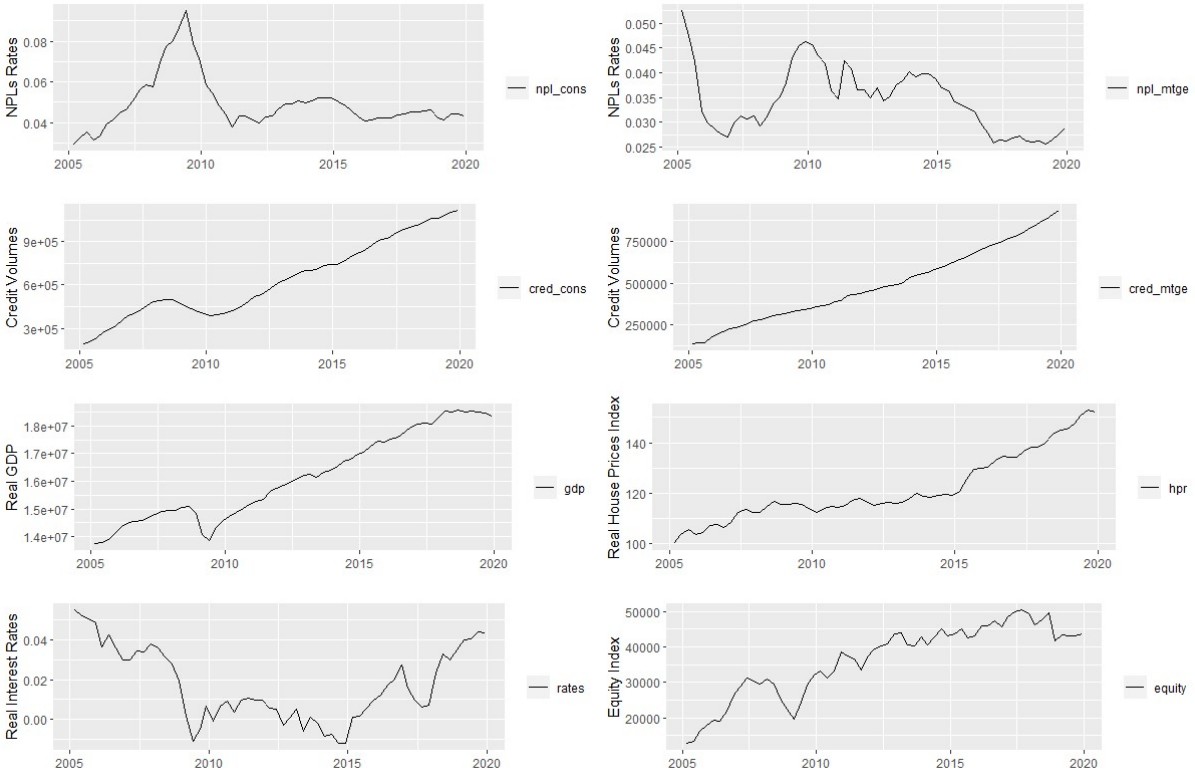
United States



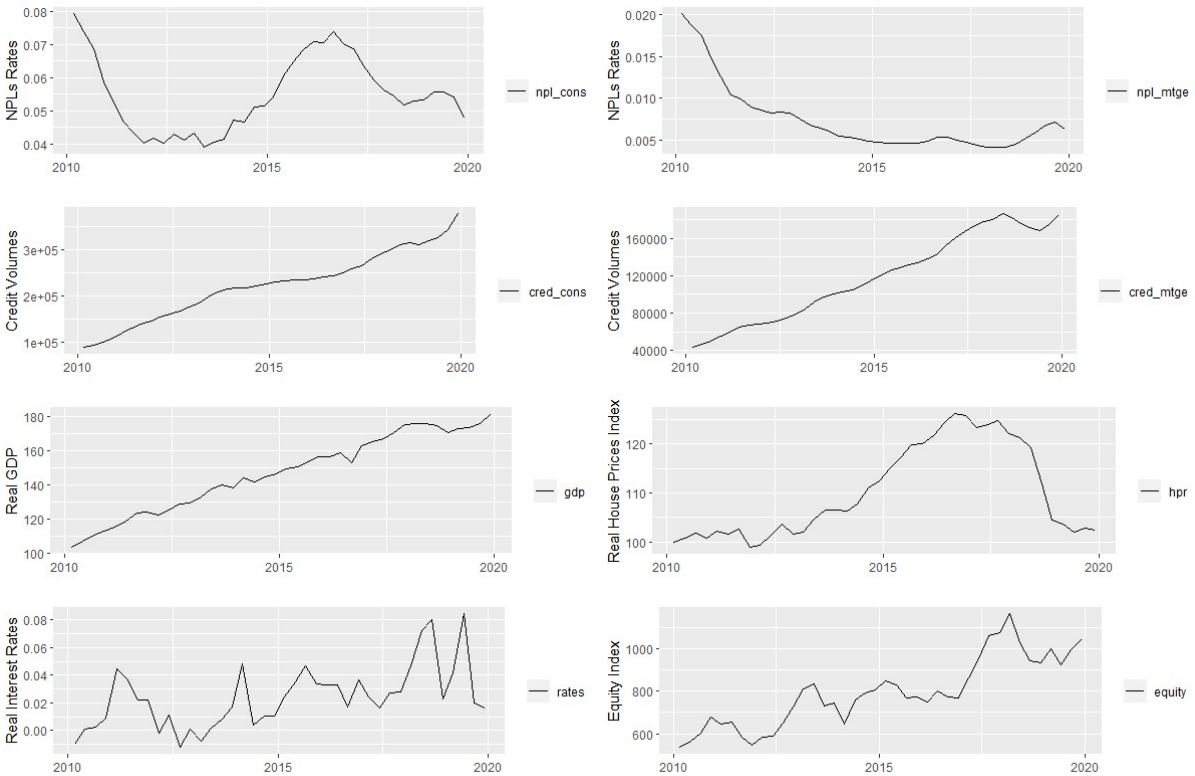
Spain



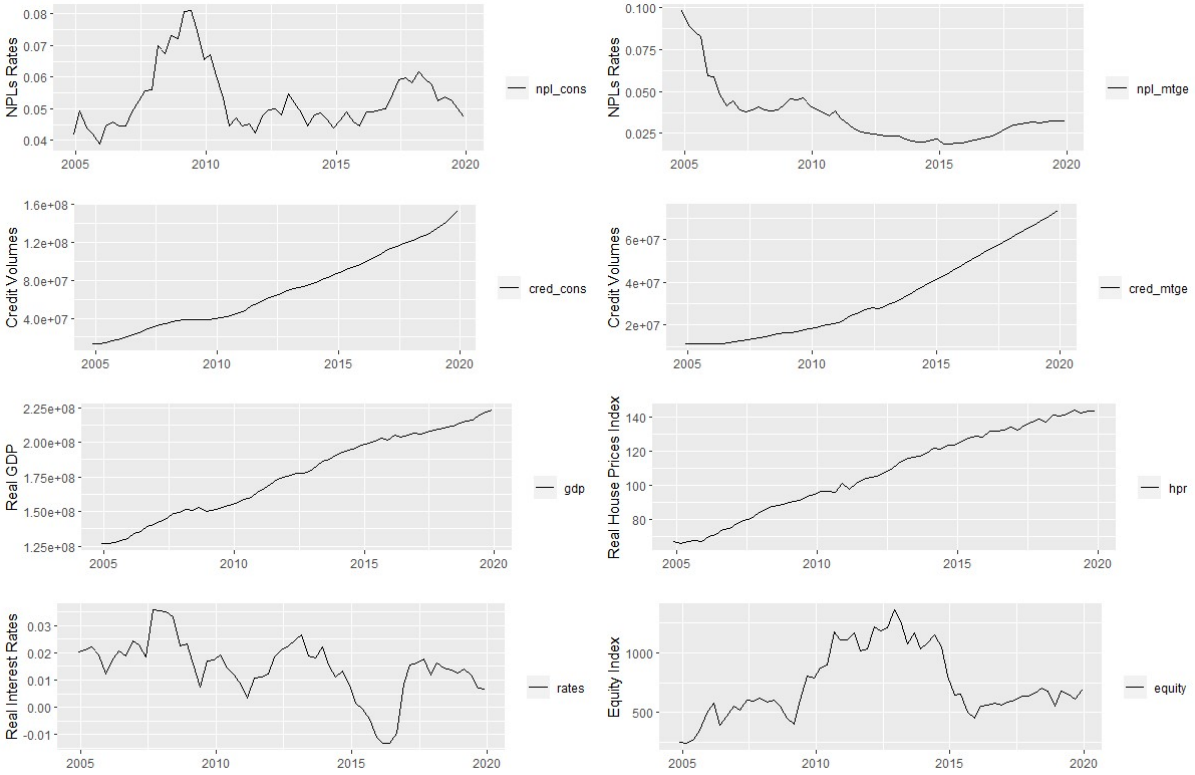
Mexico



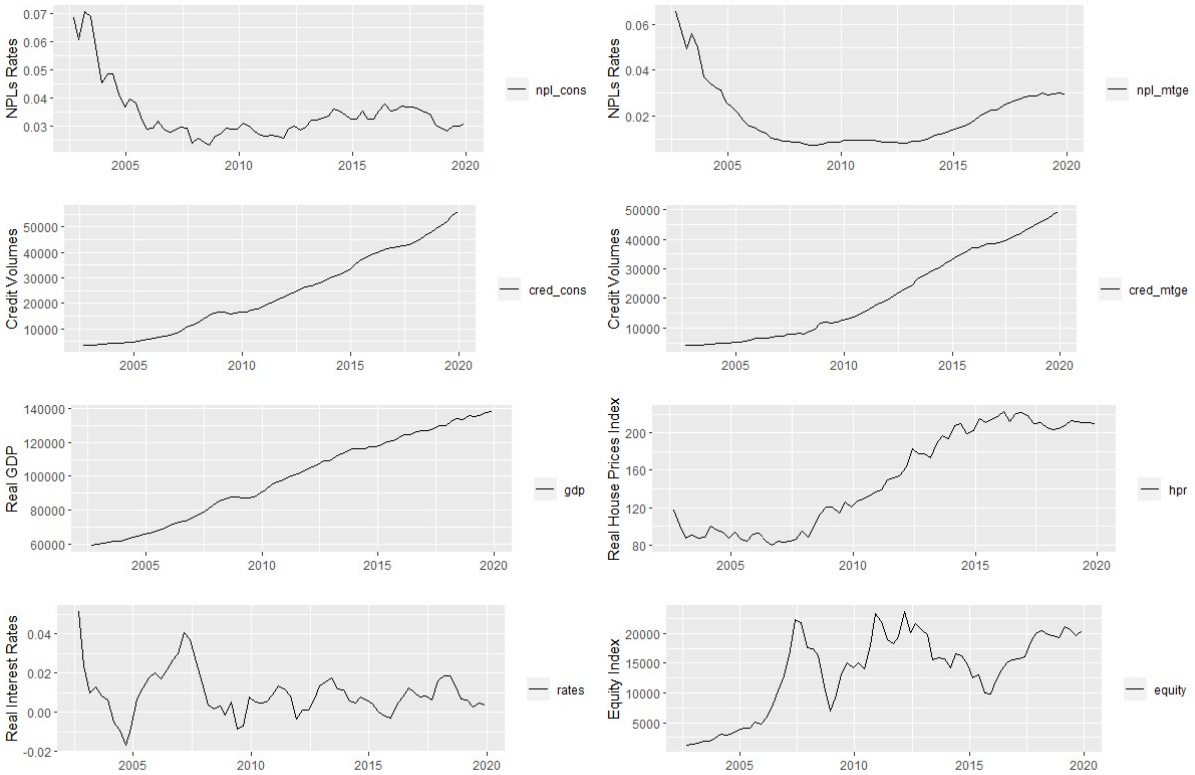
Turkey



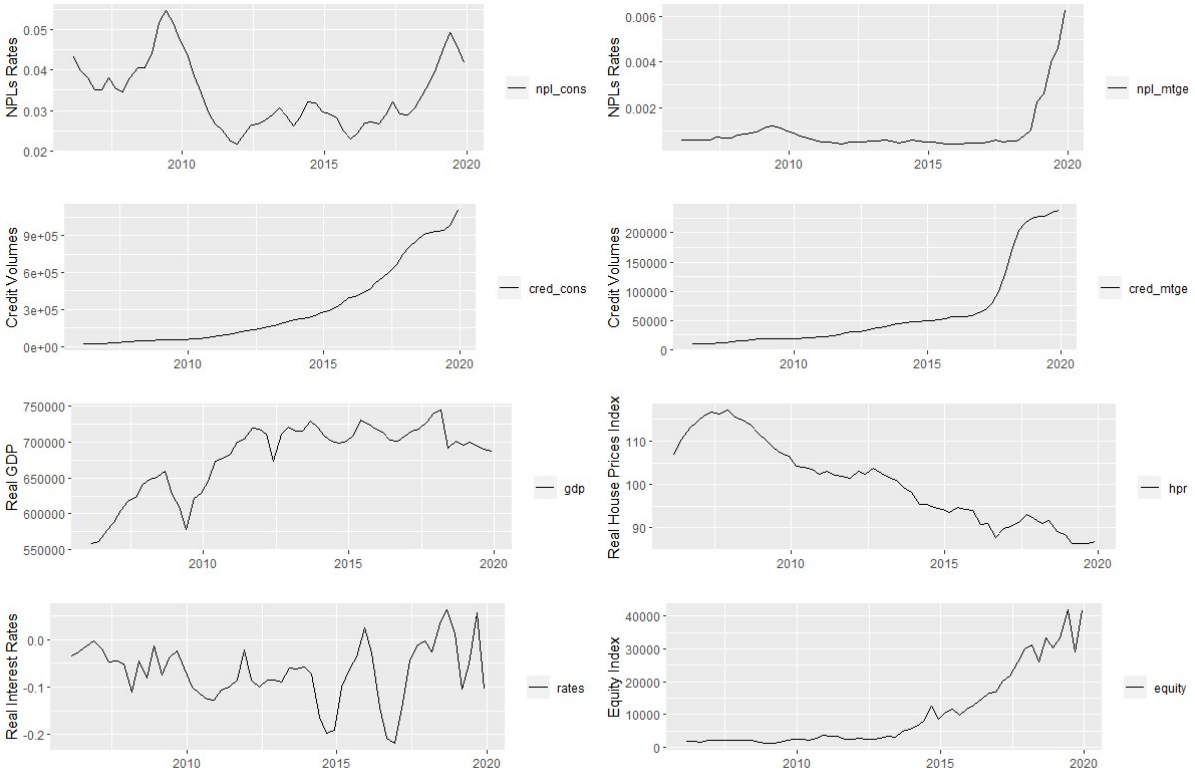
Colombia



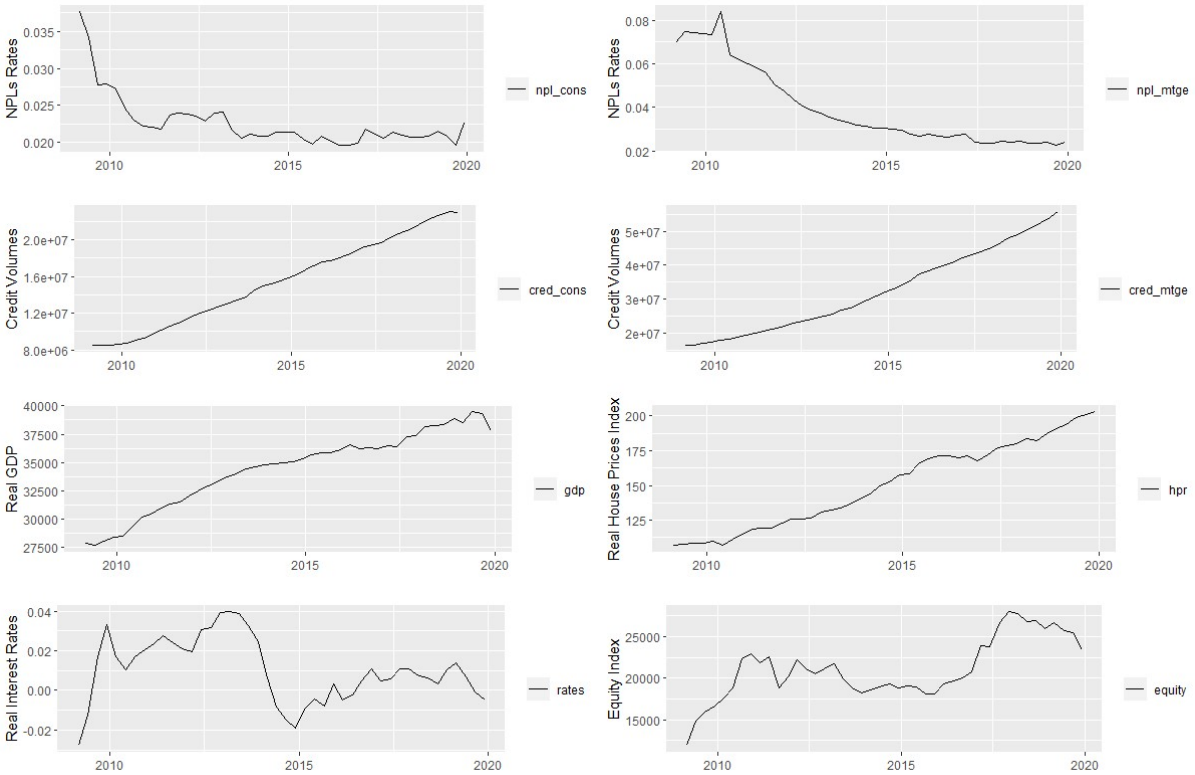
Peru



Argentina



Chile



Appendix 3. Unit root tests

Table A3.1: First Generation of Unit Root Tests

No Constant No Trend		Levin, Lin & Chu		Breitung t-stat		Im, Pesaran and Shin		ADF - Fisher		PP - Fisher Chi-square	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
log_npl_cons	level	0,02	0,508					8,16	0,944	9,50	0,891
	1st. Diff	-8,10	0,000					96,71	0,000	100,42	0,000
log_npl_mtge	level	0,38	0,650					8,41	0,936	7,91	0,952
	1st. Diff	-7,37	0,000					82,23	0,000	77,72	0,000
log_lev_cons	level	6,59	1,000					0,31	1,000	0,05	1,000
	1st. Diff	-4,05	0,000					35,45	0,003	37,93	0,002
log_lev_mtge	level	16,41	1,000					0,83	1,000	0,02	1,000
	1st. Diff	-4,46	0,000					52,47	0,000	60,60	0,000
log_gdp	level	10,08	1,000					0,30	1,000	0,16	1,000
	1st. Diff	-3,36	0,000					48,98	0,000	44,66	0,000
real rates	level	-4,74	0,000					46,39	0,000	51,75	0,000
	1st. Diff	-10,84	0,000					144,27	0,000	108,29	0,000
log_hpr	level	2,51	0,994					8,27	0,941	5,74	0,991
	1st. Diff	-4,12	0,000					47,78	0,000	47,43	0,000
log_equity	level	3,79	1,000					1,83	1,000	1,89	1,000
	1st. Diff	-8,05	0,000					88,66	0,000	86,41	0,000

Constant No Trend		Levin, Lin & Chu		Breitung t-stat		Im, Pesaran and Shin		ADF - Fisher		PP - Fisher Chi-square		Hadri	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
log_npl_cons	level	-2,68	0,004			-4,24	0,000	53,05	0,000	49,47	0,000	5,28	0,000
	1st. Diff	3,06	0,999			-4,90	0,000	57,24	0,000	57,57	0,000	1,13	0,129
log_npl_mtge	level	-3,00	0,001			-1,82	0,034	25,08	0,068	21,51	0,160	8,45	0,000
	1st. Diff	2,41	0,992			-3,25	0,001	39,07	0,001	42,68	0,000	3,62	0,000
log_lev_cons	level	-3,59	0,000			0,02	0,507	15,54	0,485	25,23	0,066	13,54	0,000
	1st. Diff	-4,34	0,000			-5,39	0,000	66,90	0,000	31,25	0,013	3,50	0,000
log_lev_mtge	level	-4,60	0,000			-0,55	0,290	27,08	0,041	39,39	0,001	14,29	0,000
	1st. Diff	-1,07	0,142			-2,71	0,003	53,13	0,000	78,07	0,000	6,42	0,000
log_gdp	level	-4,52	0,000			-1,32	0,093	23,12	0,111	26,14	0,052	14,18	0,000
	1st. Diff	0,58	0,718			-4,02	0,000	49,17	0,000	38,68	0,001	3,68	0,000
real rates	level	-1,11	0,133			-3,75	0,000	45,48	0,000	43,57	0,000	3,18	0,001
	1st. Diff	2,76	0,997			-7,28	0,000	93,33	0,000	64,37	0,000	-0,53	0,704
log_hpr	level	-3,24	0,001			0,86	0,805	21,87	0,147	8,22	0,942	10,18	0,000
	1st. Diff	-0,90	0,185			-2,75	0,003	33,12	0,007	35,63	0,003	2,82	0,002
log_equity	level	-1,93	0,027			-1,50	0,066	29,08	0,023	27,02	0,041	9,14	0,000
	1st. Diff	-0,19	0,425			-5,82	0,000	68,95	0,000	56,80	0,000	1,23	0,109

Constant & Trend		Levin, Lin & Chu		Breitung t-stat		Im, Pesaran and Shin		ADF - Fisher		PP - Fisher Chi-square		Hadri	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
log_npl_cons	level	-1,86	0,032	-0,3331	0,3695	-2,90	0,002	38,91	0,001	26,94	0,042	5,08	0,000
	1st. Diff	5,62	1,000	-1,19291	0,1165	-3,61	0,000	42,85	0,000	39,18	0,001	3,98	0,000
log_npl_mtge	level	-1,71	0,043	2,08145	0,9813	1,33	0,908	8,46	0,934	5,07	0,995	7,76	0,000
	1st. Diff	-1,23	0,110	-1,59601	0,0552	-2,88	0,002	35,37	0,004	25,09	0,068	4,53	0,000
log_lev_cons	level	-1,69	0,046	-0,09192	0,4634	-1,88	0,030	31,92	0,010	18,23	0,311	7,51	0,000
	1st. Diff	-1,93	0,027	-1,27042	0,102	-2,34	0,010	41,98	0,000	17,02	0,384	4,01	0,000
log_lev_mtge	level	-2,25	0,012	1,70536	0,9559	-0,36	0,361	41,80	0,000	50,45	0,000	9,51	0,000
	1st. Diff	0,92	0,822	0,55923	0,712	-2,53	0,006	35,57	0,003	63,85	0,000	5,78	0,000
log_gdp	level	-1,08	0,140	0,34139	0,6336	1,89	0,970	10,50	0,839	7,56	0,961	9,22	0,000
	1st. Diff	2,30	0,989	-1,59327	0,0555	-3,72	0,000	44,96	0,000	27,48	0,037	1,95	0,026
real rates	level	-1,61	0,054	-2,83951	0,0023	-3,55	0,000	42,81	0,000	34,75	0,004	2,43	0,008
	1st. Diff	4,52	1,000	-3,8257	0,0001	-7,23	0,000	84,34	0,000	42,25	0,000	0,03	0,488
log_hpr	level	-3,21	0,001	-2,35229	0,0093	0,25	0,599	22,43	0,130	14,12	0,590	6,53	0,000
	1st. Diff	-0,31	0,378	-1,97039	0,0244	-1,78	0,038	24,66	0,076	26,47	0,048	5,23	0,000
log_equity	level	-1,01	0,156	-0,47284	0,3182	-0,91	0,181	16,99	0,386	17,26	0,369	5,82	0,000
	1st. Diff	1,26	0,897	-4,71273	0	-4,69	0,000	54,73	0,000	36,73	0,002	0,99	0,160

This table computes the following five types of panel unit root tests: Levin, Lin and Chu (2002), Breitung (2000), Im, Pesaran and Shin (2003), Individual root Fisher ADF from Maddala and Wu (1999), Individual root Fisher PP fromChoi (2001), and Hadri (2000).
Sample: 3/01/1992 12/01/2019.
Automatic lag length selection based on SIC: 0 to 5.
Newey-West automatic bandwidth selection and Bartlett kernel.
Null hypothesis tests the existence of a unit root in the following tests: LLC, Breitung, IPS, ADF-Fisher and PP-Fisher. Contrary, Hadry test has a null hypothesis of no unit root .

Table A3.2: Second Generation of Unit Root Tests

Pesaran (2007) Panel Unit Root test (CIPS)

Specification in levels and without trend

Variable	lags	Zt-bar	p-value	t-bar
npl_cons	0	-0.856	0.196	.
npl_cons	1	0.310	0.622	.
npl_cons	2	0.511	0.695	.
npl_cons	3	0.552	0.710	.
npl_cons	4	-1.033	0.151	.
npl_mtge	0	1.370	0.915	.
npl_mtge	1	0.485	0.686	.
npl_mtge	2	0.518	0.698	.
npl_mtge	3	-0.068	0.473	.
npl_mtge	4	0.352	0.638	.
cred_cons	0	2.012	0.978	.
cred_cons	1	-0.011	0.496	.
cred_cons	2	-0.261	0.397	.
cred_cons	3	-0.127	0.449	.
cred_cons	4	-0.768	0.221	.
cred_mtge	0	2.338	0.990	.
cred_mtge	1	0.892	0.814	.
cred_mtge	2	2.093	0.982	.
cred_mtge	3	2.515	0.994	.
cred_mtge	4	2.209	0.986	.
gdp	0	-1.551	0.060	.
gdp	1	-1.544	0.061	.
gdp	2	-1.045	0.148	.
gdp	3	-0.826	0.205	.
gdp	4	-1.496	0.067	.
hpr	0	2.318	0.990	.
hpr	1	1.328	0.908	.
hpr	2	0.371	0.645	.
hpr	3	0.791	0.786	.
hpr	4	-1.849	0.032	.
rates	0	-3.707	0.000	.
rates	1	-5.415	0.000	.
rates	2	-4.205	0.000	.
rates	3	-3.967	0.000	.
rates	4	-2.006	0.022	.
equity	0	-1.384	0.083	.
equity	1	-1.304	0.096	.
equity	2	-1.352	0.088	.
equity	3	-1.553	0.060	.
equity	4	-1.421	0.078	.

Specification in first differences (annual) and without trend

Variable	lags	Zt-bar	p-value	t-bar
Δnpl_cons	0	-0.764	0.222	.
Δnpl_cons	1	-1.410	0.079	.
Δnpl_cons	2	-2.419	0.008	.
Δnpl_cons	3	-5.453	0.000	.
Δnpl_cons	4	0.111	0.544	.
Δnpl_mtge	0	0.318	0.625	.
Δnpl_mtge	1	-1.665	0.048	.
Δnpl_mtge	2	-1.947	0.026	.
Δnpl_mtge	3	-2.332	0.010	.
Δnpl_mtge	4	1.272	0.898	.
Δcred_cons	0	0.600	0.726	.
Δcred_cons	1	-1.898	0.029	.
Δcred_cons	2	-2.093	0.018	.
Δcred_cons	3	-1.336	0.091	.
Δcred_cons	4	1.106	0.866	.
Δcred_mtge	0	-1.168	0.121	.
Δcred_mtge	1	-3.488	0.000	.
Δcred_mtge	2	-3.180	0.001	.
Δcred_mtge	3	-1.697	0.045	.
Δcred_mtge	4	0.868	0.807	.
Δgdp	0	-3.329	0.000	.
Δgdp	1	-2.352	0.009	.
Δgdp	2	-2.160	0.015	.
Δgdp	3	-2.528	0.006	.
Δgdp	4	-0.135	0.446	.
Δhpr	0	-0.466	0.321	.
Δhpr	1	-1.898	0.029	.
Δhpr	2	-3.100	0.001	.
Δhpr	3	-2.812	0.002	.
Δhpr	4	-0.699	0.242	.
rates	0	-3.707	0.000	.
rates	1	-5.415	0.000	.
rates	2	-4.205	0.000	.
rates	3	-3.967	0.000	.
rates	4	-2.006	0.022	.
Δequity	0	-2.800	0.003	.
Δequity	1	-2.370	0.009	.
Δequity	2	-2.674	0.004	.
Δequity	3	-4.612	0.000	.
Δequity	4	-1.218	0.112	.

Specification in levels and with trend

Variable	lags	Zt-bar	p-value	t-bar
npl_cons	0	0.813	0.792	.
npl_cons	1	2.315	0.990	.
npl_cons	2	2.272	0.988	.
npl_cons	3	2.980	0.999	.
npl_cons	4	1.704	0.956	.
npl_mtge	0	5.518	1.000	.
npl_mtge	1	3.726	1.000	.
npl_mtge	2	3.340	1.000	.
npl_mtge	3	2.418	0.992	.
npl_mtge	4	2.210	0.986	.
cred_cons	0	-3.903	0.000	.
cred_cons	1	-1.898	0.029	.
cred_cons	2	-2.311	0.010	.
cred_cons	3	-2.049	0.020	.
cred_cons	4	-2.875	0.002	.
cred_mtge	0	1.198	0.884	.
cred_mtge	1	0.850	0.802	.
cred_mtge	2	1.278	0.899	.
cred_mtge	3	2.204	0.986	.
cred_mtge	4	2.300	0.989	.
gdp	0	2.059	0.980	.
gdp	1	0.837	0.799	.
gdp	2	0.905	0.817	.
gdp	3	0.238	0.594	.
gdp	4	0.444	0.672	.
hpr	0	0.216	0.585	.
hpr	1	0.332	0.630	.
hpr	2	-0.797	0.213	.
hpr	3	-0.825	0.205	.
hpr	4	-2.508	0.006	.
rates	0	-2.926	0.002	.
rates	1	-4.737	0.000	.
rates	2	-3.322	0.000	.
rates	3	-2.810	0.002	.
rates	4	-1.112	0.133	.
equity	0	0.426	0.665	.
equity	1	0.469	0.681	.
equity	2	0.733	0.768	.
equity	3	0.228	0.590	.
equity	4	0.347	0.636	.

Specification in first differences (annual) and with trend

Variable	lags	Zt-bar	p-value	t-bar
Δnpl_cons	0	0.796	0.787	.
Δnpl_cons	1	0.060	0.524	.
Δnpl_cons	2	-0.863	0.194	.
Δnpl_cons	3	-4.306	0.000	.
Δnpl_cons	4	1.943	0.974	.
Δnpl_mtge	0	1.461	0.928	.
Δnpl_mtge	1	-0.691	0.245	.
Δnpl_mtge	2	-0.962	0.168	.
Δnpl_mtge	3	-1.999	0.023	.
Δnpl_mtge	4	2.508	0.994	.
Δcred_cons	0	2.731	0.997	.
Δcred_cons	1	-0.371	0.355	.
Δcred_cons	2	-0.714	0.238	.
Δcred_cons	3	-0.063	0.475	.
Δcred_cons	4	3.454	1.000	.
Δcred_mtge	0	-0.425	0.336	.
Δcred_mtge	1	-3.750	0.000	.
Δcred_mtge	2	-3.759	0.000	.
Δcred_mtge	3	-2.341	0.010	.
Δcred_mtge	4	1.263	0.897	.
Δgdp	0	-2.336	0.010	.
Δgdp	1	-1.582	0.057	.
Δgdp	2	-1.702	0.044	.
Δgdp	3	-2.570	0.005	.
Δgdp	4	0.858	0.805	.
Δhpr	0	0.988	0.838	.
Δhpr	1	-0.516	0.303	.
Δhpr	2	-1.917	0.028	.
Δhpr	3	-2.197	0.014	.
Δhpr	4	-0.497	0.309	.
rates	0	-2.926	0.002	.
rates	1	-4.737	0.000	.
rates	2	-3.322	0.000	.
rates	3	-2.810	0.002	.
rates	4	-1.112	0.133	.
Δequity	0	-1.612	0.054	.
Δequity	1	-1.098	0.136	.
Δequity	2	-1.422	0.078	.
Δequity	3	-4.031	0.000	.
Δequity	4	0.163	0.565	.

Null for CIPS tests: series is I(1).
MW test assumes cross-section independence.
CIPS test assumes cross-section dependence is in form of a single unobserved common factor.

Conclusions

Practitioners in finance analyse a plethora of economic, financial and technical variables to determine investment decisions. They evaluate the impact of a broad set of factors on various securities' performance and take multiple choices based on the expected performance of these factors. In this sense, factor models provide a valuable tool to assist financial analysts with the identification of pervasive factors that affect a large number of securities. These factors may include macroeconomic or financial variables that gather current economic and political conditions, fundamental variables that help identify specific asset's characteristics, and even technical variables that try to pick investors sentiment.

This dissertation aims to substantiate whether macroeconomic factors and other indicators are relevant to predict both in-sample and out-of-sample assets' future performance. It investigates this practical way to forecast, focusing on two well-studied themes in financial economics and banking. First, the ability to predict the equity risk premium, and second, the macroeconomic determinants of non-performing loans (NPL) rates. In the first two chapters, a set of economic and technical metrics are examined to check whether they can forecast equity markets. Results obtained indicate that few factors show the ability to forecast out-of-sample and produce economic value for risk-averse investors. In the third chapter, similar economic factors are also explored to predict credit loan delinquencies, measured as non-performing rates, and outcomes confirm the economic cycle relevance to explain and predict delinquency rates.

Chapter 1 is entitled "Forecasting the equity risk premium in the European Monetary Union". The chapter has two main objectives. First, to analyse whether traditional economic and technical indicators show any ability to forecast the equity risk premium in the European monetary area. Second, to study whether these predictors can generate economic value for a risk-averse investor. Using extensive US datasets, the related economic literature accepts that several economic and financial indicators show in-sample forecasting power to predict stock returns. Still, there is not the same consensus when forecasting out-of-sample. Chapter 1 investigates these findings and extends the research to a less studied geographical region, the European Monetary Union.

Chapter results confirm that many economic and technical predictors can forecast EMU equity risk premiums in-sample. In particular, it finds that technical indicators display higher forecasting power than economic or valuation factors, and multivariate regressions built with principal components analysis gather relevant information from all predictors and enhance in-sample forecasting ability. However, out-of-sample results contradict in-sample ones. Technical indicators do not show out-of-sample forecasting power or the ability to create economic value for less risk-averse investors. Only a few economic factors or principal components built with these economic factors exhibit forecasting power. Just one of them, the book-to-market value, shows the ability to produce economic value for an investor with several levels of risk aversion.

Finally, some guidelines for further research in forecasting the equity risk premium are given. First, future research should investigate the forecasting power of economic and financial predictors at different time windows. Some of the indicators could forecast accurately at short periods, while others would work better at longer ones. Second, the set of variables analysed could be expanded. With the advent of "big data" techniques, practitioners find innovative ways to categorise many economic and financial variables into new measures of sentiment, risk, monetary, macroeconomic, valuation or technical variables. A third way to complement this paper would be to work with time-varying parameter models. The data-generating process for stock returns can be subject to parameter instability. Models that assume that estimation parameters can take different values as the economy switches between economic regimes could improve models predictability power.

Chapter 2, entitled "Forecasting the European Monetary Union equity risk premium with regression trees ", expands on chapter 1 by investigating whether a popular machine learning technique, classification and regression trees (CART), can contribute to improve EMU equity risk premium forecasts. The literature covering the capacity of machine learning algorithms to enhance equity risk premiums forecasts is still not extensive, and this essay aims to contribute to it. The work implements three ensemble methods (bagging, random forests and boosting) and investigates whether these algorithms can improve the ability to forecast out-of-sample the equity risk premium.

Results indicate that regression trees do not show more forecasting power than a naïve benchmark or univariate regressions over individual predictors. These outcomes contradict those obtained by Gu et al. (2020), who find in a US dataset that regression trees offer high

out-of-sample forecasting power for an extensive data sample, including 30.000 individual stocks, over 60 years, and more than 900 predictors. These differences between the studies' outcomes might suggest that regression trees and other machine learning techniques can be a helpful forecasting tool in richer and transversal datasets environments. In contrast, smaller dimension samples might be better analysed with traditional parametric analysis.

Complementarily, chapter 2 conducts a simple asset allocation exercise, using Brandt and Santa-Clara (2006) conditional portfolio choice to study whether regression tree algorithms can generate economic value for an investor. The optimisation results are mixed because portfolio performance indicators do not point in the same direction. While tree algorithms develop higher relative certainty equivalent return (CER) for regression tree strategies, the benchmark's Sharpe ratio exhibits a greater value than the trees' ones.

We believe these results are of interest for many practitioners and academics already implementing these new machine learning techniques in the finance field. Our main findings suggest that regression trees do not seem to perform better than simple parametric models in low dimensional datasets. Therefore, further research may be needed with higher dimension samples, where traditional parametric models struggle to find consistent patterns, while machine learning techniques aggregate and detect multiple relationships easily.

The third chapter also explores similar economic factors, but this time predicts credit loan delinquencies, measured as non-performing rates. To do so, traditionally, aggregated portfolio data has been employed, and panels with a large number of cross-sections used. However, when researchers or practitioners want to work with more disaggregated portfolios, such data does not exist or it is costly to collect.

Chapter 3, entitled "Macro determinants of non-performing loans: A comparative analysis between consumer and mortgage loans", arranges an intermediate and heterogenous panel. The dataset includes eight developed and emerging economies: United States, Spain, Mexico, Turkey, Colombia, Peru, Argentina and Chile. It builds two unbalanced dynamic panels, one for consumer loans and the other for mortgages, and explores the ability of few economic factors to explain and forecast these two portfolios delinquency rates. The principal objective of the chapter is to confirm previous literature results but dealing with more disaggregated data and for a panel with a short number of transversal observations and more extensive time data than cross-sections. Practitioners often have to model and predict with not enough data, as it is

the case with non-performing loans, and this chapter aims to contribute to the scarce literature covering these situations.

Results obtained are in line with related literature, and macroeconomic factors are relevant determinants of non-performing loans (NPL). Delinquency rates show a persistent nature, past credit growth has a positive impact on NPLs, and an acceleration in real GDP, real house prices and equity markets, along with lower funding costs, lead to lower NPLs. Estimations also confirm that consumer loans and mortgages display very similar sensibility to the economic cycles. Yet, the elasticities are slightly different and invite to estimate these two portfolios independently. In addition, in-sample estimations also showed that coefficients obtained through FE-WG and instrumental variables look very similar, making us wonder if the sample T is long enough to remove FE-WG bias in these dynamic panels. Finally, the out-of-sample forecast confirmed that most models tend to beat the benchmark and economic factors play an essential role in forecasting non-performing loans.

Chapter 3 findings could be of relevance for academics and practitioners in two ways. First, outcomes contribute to the extensive literature exploring the macroeconomic determinants of non-performing loans. It finds similar results but focuses on intermediate panels, “large T, small N” panels, where the literature is not abundant, and practitioners sometimes have to work with. Second, the essay contributes to the debate of whether to pool or not pool in intermediate-sized panels and if static econometric techniques are suitable for dynamic panels when the time dimension is large.

Looking ahead, future research to complement and improve the present study should research further in the lag structure of the autoregressive distributed panel models and investigate potential causality relations. In this line, estimating a PVAR system could offer a robustness check of the panel regression results. Moreover, as richer datasets appear at a disaggregated portfolio level, macro panel techniques should also be implemented to improve forecasts and confirm the obtained results.